

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

**ADVANCEMENTS IN SEISMIC TOMOGRAPHY
WITH APPLICATION TO
TUNNEL DETECTION AND VOLCANO IMAGING**

**A
THESIS**

Presented to the Faculty
of the University of Alaska Fairbanks
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

By James Doyle Clippard, B.S.

Fairbanks, Alaska

1998

UMI Number: 9903663

UMI Microform 9903663
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

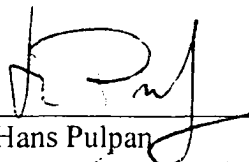
UMI
300 North Zeeb Road
Ann Arbor, MI 48103

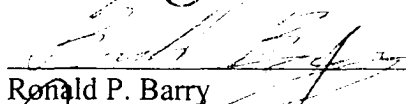
**ADVANCEMENTS IN SEISMIC TOMOGRAPHY
WITH APPLICATION TO
TUNNEL DETECTION AND VOLCANO IMAGING**

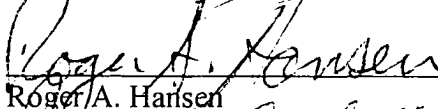
By

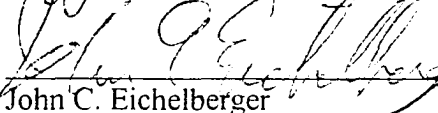
James D. Clippard

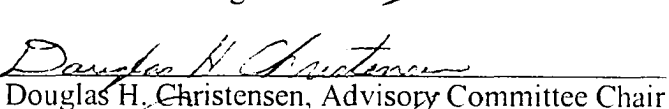
RECOMMENDED:



Hans Pulpan


Ronald P. Barry

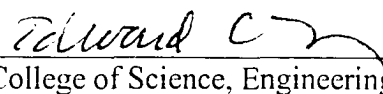

Roger A. Hansen

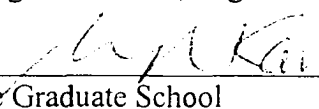

John C. Eichelberger

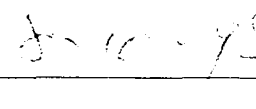

Douglas H. Christensen, Advisory Committee Chair


Paul W. Layer, Department Head

APPROVED:


Dean, College of Science, Engineering, and Mathematics


Dean of the Graduate School


Date

ABSTRACT

Practical geotomography is an inverse problem with no unique solution. *A priori* information must be imposed for a stable solution to exist. Commonly used types of *a priori* information smooth and attenuate anomalies, resulting in ‘blurred’ tomographic images. Small or discrete anomalies, such as tunnels, magma conduits, or buried channels are extremely difficult imaging objectives.

Composite distribution inversion (CDI) is introduced as a theory seeking physically simple, rather than distributionally simple, solutions of non-unique problems. Parameters are assumed to be members of a composite population, including both well-known and anomalous components. Discrete and large amplitude anomalies are allowed, while a well-conditioned inverse is maintained.

Tunnel detection is demonstrated using CDI tomography and data collected near the northern border of South Korea. Accurate source and receiver location information is necessary. Borehole deviation corrections are estimated by minimizing the difference between empirical distributions of apparent parameter values as a function of location correction. Improved images result.

Traveltime computation and raytracing are the most computationally intensive components of seismic tomography when imaging structurally complex media. Efficient, accurate, and robust raytracing is possible by first recovering approximate raypaths from traveltime fields, and then refining the raypaths to a desired accuracy level. Dynamically binned queuing is introduced. The approach optimizes graph-theoretic traveltime

computation costs. Pseudo-bending is modified to efficiently refine raypaths in general media.

Hypocentral location density functions and relative phase arrival population analysis are used to investigate the Spring, 1996, earthquake swarm at Akutan Volcano, Alaska. The main swarm is postulated to have been associated with a 0.2 km^3 intrusion at a depth of less than four kilometers. Decay sequence seismicity is postulated to be a passive response to the stress transient caused by the intrusion.

Tomograms are computed for Mt. Spurr, Augustine, and Redoubt Volcanoes, Alaska. Relatively large amplitude, shallow anomalies explain most of the traveltime residual. No large amplitude anomalies are found at depth, and no magma storage areas are imaged. A large amplitude low-velocity anomaly is coincident with a previously proposed geothermal region on the southeast flank of Mt. Spurr. Mt. St. Augustine is found to have a high velocity core.

Table of Contents

APPROVED:	2
ABSTRACT	3
TABLE OF CONTENTS	5
LIST OF FIGURES.....	8
LIST OF TABLES.....	10
ACKNOWLEDGEMENTS	11
CHAPTER 1	15
<i>INTRODUCTION TO THE DISSERTATION</i>	<i>15</i>
OBJECTIVE	15
INTRODUCTION TO THE CHAPTERS	16
SIGNIFICANCE OF THE WORK	18
CONCLUSIONS	20
RECOMMENDATIONS FOR FUTURE INVESTIGATIONS	21
MATERIAL CONTRIBUTORS TO THIS WORK	22
REFERENCES	23
CHAPTER 2	24
<i>COMPOSITE DISTRIBUTION INVERSION APPLIED TO CROSSHOLE TOMOGRAPHY</i>	<i>24</i>
ABSTRACT.....	24
INTRODUCTION.....	25
METHODS.....	27
<i>Problem formulation</i>	<i>27</i>
<i>Appropriate a priori information</i>	<i>30</i>
<i>Implementation</i>	<i>35</i>
RESULTS.....	36
<i>Evaluation methods</i>	<i>36</i>
<i>Field data application</i>	<i>40</i>
DISCUSSION AND CONCLUSIONS	42
ACKNOWLEDGEMENTS.....	44
REFERENCES	45
CHAPTER 3	58
<i>TUNNEL DETECTION USING TRAVELTIME AND ATTENUATION CDI TOMOGRAPHY IN SOUTH KOREA</i>	<i>58</i>

ABSTRACT.....	58
INTRODUCTION.....	59
METHODS.....	61
<i>Traveltime tomography and empirical corrections</i>	64
<i>Attenuation tomography and empirical corrections</i>	67
DATA ACQUISITION.....	73
DATA ANALYSIS AND RESULTS.....	74
CONCLUSIONS	79
ACKNOWLEDGEMENTS.....	80
REFERENCES	80
CHAPTER 4	93
<i>3D SEISMIC RAYTRACING USING GRAPH THEORETIC TRAVELTIME FIELDS AND RAY REFINEMENT</i>	93
ABSTRACT.....	93
INTRODUCTION.....	94
<i>Motivation and General Overview</i>	94
<i>Design Philosophy</i>	98
TRAVELTIME FIELD ESTIMATION	100
<i>General Algorithm Selection</i>	100
<i>Graph Theory Traveltime Field Calculation</i>	102
GRAPH THEORY ALGORITHM OPTIMIZATION.....	107
<i>Ray Segment Traveltime Calculation</i>	107
<i>Efficient Forward-Star Design Details</i>	109
<i>Efficient Queueing</i>	111
PSEUDO-BENDING RAYPATH REFINEMENT	115
<i>Basic Pseudo-Bending</i>	118
<i>Performance Enhancement</i>	120
<i>Strongly Non-Linear Models</i>	122
<i>Snell's Law, Constant Velocity Regions, and Topography</i>	126
RESULTS.....	129
<i>Performance with a Constant Gradient Model</i>	129
<i>Performance with an Exponential Velocity Model</i>	133
<i>Performance with a Heterogeneous Grid Model</i>	133
CONCLUSIONS.....	134
ACKNOWLEDGEMENTS.....	135
REFERENCES	137
CHAPTER 5	156
<i>ANALYSIS OF THE SPRING 1996 EARTHQUAKE SWARM AT AKUTAN VOLCANO, ALASKA</i>	156
ABSTRACT.....	156
INTRODUCTION.....	157
SWARM DATA AND ANALYSIS.....	160
<i>Chronology of the Akutan Swarm, Available Data, and Operational Objectives</i>	160
<i>Early Episode Magnitudes, Seismic Energy, and b-Values</i>	162
<i>Decay Sequence Data</i>	167
<i>Omori Law p-Values for the Decay Sequence</i>	168
<i>Temporal Stability of the Source Concentration Area</i>	171
<i>Akutan Velocity Model</i>	174
<i>Hypocenter Estimation: Issues, Requirements, and Methods</i>	179

<i>Akutan Hypocenter Estimates</i>	186
DISCUSSION.....	189
CONCLUSIONS	191
ACKNOWLEDGEMENTS.....	192
REFERENCES	193
CHAPTER 6	220
<i>SEISMIC TOMOGRAPHY FOR VOLCANO IMAGING: DESIGN, DEVELOPMENT, AND APPLICATION TO REDOUBT, MT. SPURR, AND AUGUSTINE VOLCANOES, ALASKA</i>	220
ABSTRACT.....	220
INTRODUCTION.....	221
METHODS.....	224
<i>Data Requirements and Selection</i>	226
<i>Traveltime Calculation and Velocity Model Basis Function Selection</i>	229
<i>A Priori Information</i>	233
<i>Parameterization of the Primary Free Parameters</i>	235
<i>Estimation of Earthquake Source Locations</i>	239
<i>Coordinate System</i>	241
<i>Solving the Linear System</i>	242
DATA	243
<i>Redoubt Data</i>	245
<i>Spurr Data</i>	246
<i>Augustine Data</i>	246
RESULTS.....	248
<i>Redoubt</i>	248
<i>Spurr</i>	250
<i>Augustine</i>	251
DISCUSSION.....	253
<i>Redoubt</i>	254
<i>Spurr</i>	255
<i>Augustine</i>	256
CONCLUSIONS	257
ACKNOWLEDGEMENTS.....	258
REFERENCES	259

List of Figures

FIGURE 1: COMPARISON OF CDI PENALTY FUNCTION WITH DAMPED LEAST SQUARES.....	49
FIGURE 2: RAYPATHS USED FOR POINT SPREAD FUNCTION EXAMPLES.	50
FIGURE 3: POINT SPREAD FUNCTIONS.....	51
FIGURE 4: BRICK ANOMALY IMAGING COMPARISON.	52
FIGURE 5: BLOB ANOMALY IMAGING COMPARISON.....	53
FIGURE 6: PLUS ANOMALY IMAGING COMPARISON.	54
FIGURE 7: AZIMUTHS USED FOR CROSSHOLE DATA ACQUISITION.....	55
FIGURE 8: TOMOGRAMS BASED ON CROSSHOLE FIELD DATA.....	56
FIGURE 9: COMPARISON OF ESTIMATED VELOCITY DISTRIBUTIONS.....	57
FIGURE 10: DIFFERENCES IN AREA SAMPLED BY DIFFERENT ACQUISITION ANGLES.....	83
FIGURE 11: DIAGRAM OF EXPERIMENTAL OBJECTIVE.....	84
FIGURE 12: ARC-DISCHARGE SOURCE USED FOR THE EXPERIMENT.....	85
FIGURE 13: CROSSHOLE SAMPLING ANGLES.....	86
FIGURE 14: DAMPED LEAST-SQUARE TRAVELTIME TOMOGRAMS WITHOUT EMPIRICAL CORRECTIONS.....	87
FIGURE 15: DAMPED LEAST-SQUARE ATTENUATION TOMOGRAM WITHOUT EMPIRICAL CORRECTIONS.....	88
FIGURE 16: STABILITY OF BOREHOLE DEVIATION CORRECTION ESTIMATES INDICATED BY A NORMALIZED DISPERSION MEASURE.....	89
FIGURE 17: MEDIAN VELOCITY AS A FUNCTION OF SAMPLING ANGLE.....	90
FIGURE 18: FINAL CDI ATTENUATION TOMOGRAMS.....	91
FIGURE 19: COMPARISON OF UNDERDAMPED LEAST SQUARE AND CDI TRAVELTIME TOMOGRAMS.....	92
FIGURE 20: NODES CORRESPONDING TO EDGES OF A REDUCED, 3-D, FORWARD-STAR, IN THE X-Y PLANE.....	145
FIGURE 21: COMPARISON OF EDGE COUNTS FOR REDUCED AND UNREDUCED FORWARD-STAR.....	146
FIGURE 22: NUMBER OF REQUIRED QUEUEING OPERATIONS VERSUS TRAVELTIME INTERVAL FOR VARIOUS PROPOSED QUEUES.	147
FIGURE 23: MAXIMUM ERROR AND COMPUTATION COST FOR VARIOUS RELAXATION SCHEDULES USING ONE- THOUSAND RAYS AND A CONSTANT GRADIENT VELOCITY MODEL.....	148
FIGURE 24: THIRTEEN RAYS IN AN EXPONENTIAL VELOCITY MODEL.....	149
FIGURE 25: ERROR AND COMPUTATION COST FOR VARIOUS RELAXATION SCHEDULES USING THIRTEEN RAYS AND AN EXPONENTIALLY INCREASING VELOCITY MODEL.....	150
FIGURE 26: TRAVELTIME PERCENTAGE ERRORS FOR 1000 RANDOM LOCATIONS IN A CONSTANT GRADIENT MODEL.....	151
FIGURE 27: MAXIMUM ABSOLUTE ERRORS FOR 1000 RANDOM LOCATIONS IN THE CONSTANT GRADIENT MODEL.....	152
FIGURE 28: COMPUTING COST AND NUMBER OF NODES AS A FUNCTION OF GRID SPACING OF THE (10 KM X 10 KM X 5 KM) STUDY VOLUME.	153
FIGURE 29: COMPUTATIONAL COST AND MAXIMUM ERROR FOR 100,000 RANDOMLY DISTRIBUTED RAYS.....	154
FIGURE 30: RAYPATHS IN A HETEROGENEOUS, GRID-BASED, MODEL.....	155
FIGURE 31: LOCATION OF AKUTAN ISLAND AND STATION DTN.	201
FIGURE 32: AKUTAN EVENT COUNTS PER HOUR OBSERVED AT STATION DTN.....	202
FIGURE 33: MAP OF AKUTAN ISLAND SHOWING LOCATION OF AKUTAN SEISMIC STATIONS, CALDERA, CITY, AND DECAY SEQUENCE EPICENTERS.	203
FIGURE 34: ESTIMATED DURATION MAGNITUDE VERSUS AEIC OR NEIC CATALOG MAGNITUDE.....	204
FIGURE 35: EARTHQUAKE MAGNITUDE VERSUS TIME FOR THE FIRST TWO EPISODES OF THE AKUTAN SWARM.	205
FIGURE 36: CUMULATIVE MOMENT VERSUS DATE FOR THE FIRST TWO EPISODES OF SEISMICITY.....	206
FIGURE 37: CUMULATIVE ENERGY RELEASE VERSUS ELAPSED TIME AFTER EPISODE INCEPTION FOR THE FIRST TWO EPISODES.....	207
FIGURE 38: MAGNITUDE VERSUS LOG₁₀ (RANK).....	208

FIGURE 39: COMPARISON OF OBSERVED EARTHQUAKE COUNTS WITH THOSE PREDICTED BY THE MODIFIED OMORI LAW.....	209
FIGURE 40: DECAY SEQUENCE MAGNITUDE VERSUS TIME.....	210
FIGURE 41: <i>P</i> -WAVE ARRIVAL TIME AT AK1 AND AK3 RELATIVE TO AK2 VERSUS DATE AND TIME OF OBSERVATION.....	211
FIGURE 42: <i>P</i> -WAVE ARRIVAL TIME AT AK4 RELATIVE TO AK2 VERSUS DATE AND TIME OF OBSERVATION.....	212
FIGURE 43: <i>P</i> -WAVE ARRIVAL TIME AT AK5 RELATIVE TO AK2 VERSUS DATE AND TIME OF OBSERVATION.....	213
FIGURE 44: RATIO OF <i>P</i> -WAVE TO <i>S</i> -WAVE TRAVELTIME DIFFERENCES USED FOR ESTIMATING V_p/V_s	214
FIGURE 45: TRAVELTIME RESIDUALS VERSUS DISTANCE FROM HYPOCENTER.....	215
FIGURE 46: TARANTOLA-VALETTE L_2 NORM EPICENTRAL DENSITY FUNCTION AND 432 MAXIMUM LIKELIHOOD EPICENTRAL ESTIMATES IN THE DETAIL AREA.....	216
FIGURE 47: TARANTOLA-VALETTE L_2 NORM HYPOCENTRAL LOCATION DENSITY FUNCTION IN THE VERTICAL PLANE A-A' FOR DOMINANT MODE ARRIVALS ASSUMING A PICKING ERROR STANDARD DEVIATION OF 0.1 SECOND AT ALL LOCATIONS.....	217
FIGURE 48: L_1 NORM HYPOCENTRAL LOCATION DENSITY FUNCTION IN THE VERTICAL PLANE A-A' FOR DOMINANT MODE ARRIVALS ASSUMING A PICKING ERROR STANDARD DEVIATION OF 0.1 SECOND.....	218
FIGURE 49: L_2 NORM EPICENTRAL DENSITY FUNCTION FOR SECONDARY MODE ARRIVALS AND 432 MAXIMUM LIKELIHOOD EPICENTRAL ESTIMATES IN THE DETAIL AREA.....	219
FIGURE 50: LOCATION OF REDOUBT, SPURR, AND AUGUSTINE VOLCANOES.....	268
FIGURE 51: REDOUBT VOLCANO STUDY AREA.....	269
FIGURE 52: INITIAL EPICENTER ESTIMATES AND SHOT LOCATIONS USED FOR TOMOGRAPHIC IMAGING OF REDOUBT VOLCANO.....	270
FIGURE 53: ONE DIMENSIONAL REDOUBT VELOCITY MODELS.....	271
FIGURE 54: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR REDOUBT VOLCANO AT 2 KM AND 1.2 KM ABOVE SEA LEVEL.....	272
FIGURE 55: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR REDOUBT VOLCANO AT 0.4 KM ABOVE SEA LEVEL AND 0.4 KM BELOW SEA LEVEL.....	273
FIGURE 56: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR REDOUBT VOLCANO AT 1.2 KM AND 2.0 KM BELOW SEA LEVEL.....	274
FIGURE 57: HORIZONTAL SLICES OF <i>S</i> WAVE VELOCITY PERTURBATION FOR REDOUBT VOLCANO AT 2 KM AND 1.2 KM ABOVE SEA LEVEL.....	275
FIGURE 58: HORIZONTAL SLICES OF <i>S</i> WAVE VELOCITY PERTURBATION FOR REDOUBT VOLCANO AT 0.4 KM ABOVE SEA LEVEL AND 0.4 KM BELOW SEA LEVEL.....	276
FIGURE 59: HORIZONTAL SLICES OF <i>S</i> WAVE VELOCITY PERTURBATION FOR REDOUBT VOLCANO AT 1.2 KM ABOVE SEA LEVEL AND 2 KM BELOW SEA LEVEL.....	277
FIGURE 60: HORIZONTAL SLICES OF V_p/V_s ESTIMATE FOR REDOUBT VOLCANO AT 2 KM AND 1.2 KM ABOVE SEA LEVEL.....	278
FIGURE 61: HORIZONTAL SLICES OF V_p/V_s ESTIMATE FOR REDOUBT VOLCANO AT 0.4 KM ABOVE SEA LEVEL AND 0.4 KM BELOW SEA LEVEL.....	279
FIGURE 62: HORIZONTAL SLICES OF V_p/V_s FOR REDOUBT VOLCANO AT 1.2 KM AND 2.0 KM BELOW SEA LEVEL.....	280
FIGURE 63: MT. SPURR STUDY AREA.....	281
FIGURE 64: ONE DIMENSIONAL MT. SPURR VELOCITY MODELS.....	282
FIGURE 65: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR MT. SPURR AT 2.0 AND 1.0 KM ABOVE SEA LEVEL.....	283
FIGURE 66: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR MT. SPURR AT SEA LEVEL AND 1 KM BELOW SEA LEVEL.....	284
FIGURE 67: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR MT. SPURR AT 2 KM AND 3 KM BELOW SEA LEVEL.....	285

FIGURE 68: VERTICAL CROSS-SECTION OF <i>P</i> WAVE VELOCITY PERTURBATION FOR MT. SPURR ALONG PROFILES A-A' AND B-B' SHOWN IN FIGURE 63. A) A-A'; B) B-B'.....	286
FIGURE 69: FINAL RELOCATED HYPOCENTERS ACTUALLY USED AS SOURCES FOR TOMOGRAPHIC IMAGING AT MT. SPURR.....	287
FIGURE 70: AUGUSTINE VOLCANO STUDY AREA.....	288
FIGURE 71: AUGUSTINE VOLCANO VELOCITY MODEL AREA.	289
FIGURE 72: ONE DIMENSIONAL AUGUSTINE VELOCITY MODELS.	290
FIGURE 73: AUGUSTINE VOLCANO VELOCITY MODEL PROPOSED BY KIENLE ET AL. (1977).....	291
FIGURE 74: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR AUGUSTINE VOLCANO AT 1 KM AND 0.75 KM ABOVE SEA LEVEL.....	292
FIGURE 75: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR AUGUSTINE VOLCANO AT 0.5 KM AND 0.25 KM ABOVE SEA LEVEL.....	293
FIGURE 76: HORIZONTAL SLICES OF <i>P</i> WAVE VELOCITY PERTURBATION FOR AUGUSTINE VOLCANO AT SEA LEVEL AND 0.25 KM BELOW SEA LEVEL.....	294

List of Tables

TABLE 1: COMPARISON OF PSEUDO-BENDING PERFORMANCE USING LINEAR SLOWNESS VERSUS LINEAR VELOCITY SEGMENT TRAVELTIME ESTIMATION.	143
TABLE 2: COMPARISON OF TRAVELTIME FIELD CALCULATION COSTS AND ASSOCIATED MAXIMUM ERRORS FOR VARIOUS FORWARD-STAR LEVELS AND GRID NODE SPACINGS.	144

ACKNOWLEDGEMENTS

I have had the benefit of advice and support from many people while pursuing the projects that culminated in this thesis. I haven't the time or inclination to fairly acknowledge everyone, but I would like to thank a few individuals that come to mind at the moment.

First, I thank my advisor, Douglas Christensen, for unflagging support and creative input. Although I doubt that anyone would understand my perseverance through the inverse forest, Doug tolerated and guided my efforts. Perhaps he didn't fully guide, but only because I was so difficult to steer. The academic world would undoubtedly be a happier place if all graduate students could have someone like Doug as an advisor. It doesn't get any better.

Next, my committee members deserve thanks. Hans Pulpan always provided positive reinforcement, excellent intuition, and a seasoned understanding of the philosophy behind my quest. Keith Echelmeyer's dynamic approach to both field and office work provided continuous inspiration. Roger Hansen most generously stepped in to fill the void left by Echelmeyer's sabbatical during my final semester. Ron Barry made me think of significant things, interesting units of measure, and p-values. John Eichelberger provided an opportunity for me to participate in the operations of the Alaska Volcano Observatory. He also filled a very large committee void left by the death of Juergen Kienle. Juergen and I shared a few common dreams, academic and otherwise. We bonded both at the *Ivory Tower* and in pursuit of data, truth, and beauty. I will always be haunted by his absence.

Though not a committee member, Richard Rechtien was my first geophysics professor and has provided me with academic guidance since 1981. Data he provided allowed demonstration of the ‘real-world’ effectiveness of composite distribution inversion as applied to seismic tomography. I am grateful for the opportunity to co-publish with Dick while at UAF. Gerald Rupert should be also be mentioned as the single individual most responsible for my initial studies in geophysics. As a direct consequence I can comfortably state that “life’s been good to me so far”.

John Power provided picks of 1986 Augustine pre-eruption seismic events, vast amounts of reference material, Akutan paper records, and his measurements of Akutan coda duration, which provided a much needed reality check on my interpretation. He was briefly an office-mate and I am most grateful for the role he played in securing my position with AVO. John defines my concept of ‘USGS seismologist’.

Neither David Stone nor Keith Runcorn were on my committee, but I was most influenced by their discourse. I managed to gather Dave’s perspective more or less once a week for several years, and Keith’s more frequently than that during Fall semesters. If I could salvage anything lost while at UAF it would be ‘snippets’ of conversation with David or Keith that I must have forgotten.

Many students provided support beyond superlative description. I feel most fortunate to have been a part of the Geophysical Institute seismology program during a period of critical mass. Hilary Fletcher and Scott Stihler both exceeded all my fantasies of what an ‘office-mate’ might be. Cheryl Searcy and Dapeng Zhao shared not only my office, but an interest in tomography. John Benoit and Art Jolly helped me appreciate

things volcanic. The entire Augustine field crew deserves an accolade for a job very nearly well done, at a cost of roughly $\frac{1}{20}$ of what I thought was needed.

The support Mitch Robinson provided is almost indescribable. He administered the amazing growth of the seismology SPARC cluster. I can honestly say that my thesis work could not have been completed without the system and people skills of Mitch. Roger Hansen also deserves applause in the computing area. He understands what ‘technical advantage’ means, and has applied the full-court press for seismo-computing technology enhancement. He also gave me more CPU cycles from his personal machine than most people have the option of offering. Roger, I owe you one.

Others about the institute that were most supportive include Steve Estes, John “Benny” Benedetto, Kevin Abinett, Kent Lindquist, Gil Mull, Max Wyss, Stefan Wiemer, and Steve McNutt. Steve provided candid input, and a place to live while porting 100,000+ lines of C++ code to a newly installed Gnu G++ compiler. It was during this time that I introduced my family to Fairbanks living. Steve deserves credit for the taste of Fairbanks that still lingers in their hearts and minds.

All members of the GI business and operations office were a joy to work with. They continue to provide solid evidence that it is possible for a government / academic administrative office to actually support the operations of the unit supporting them. Bob Grove and Penelope Noecker deserve special mention in this arena. Operational funding is also very important, and support from the NSF, Alaska Volcano Observatory, Alaska Geophysical Society, Exxon, and the USGS is acknowledged in each chapter.

Prior to arriving at the UAF Geophysical Institute, I already had many ideas in place that contributed towards my pursuits at UAF. I didn't know it when I arrived at the GI, but I now realize that my most enduring pre-GI geophysical thoughts were primarily motivated by interaction with Ralph and Richard Simon, Carey Russell, Mike Glinsky, and Joel Hayden. Ralph, who invented the first color camera that went to the moon, was quite a business partner. His achievements certainly enhanced my appreciation of technology oriented research. Richard provided a Harvard perspective that solidly affirmed my belief in the GI program. Carey helped me understand the acid bath, and how the bath relates to geophysical inverse theory. While at Shell, Carey and I formulated the Law of Constant Complexity. The formulation of this law remains a significant achievement, and I have referred to it on more than one occasion while problem-solving at UAF. Mike and I thought together long ago, his influence and input is commensurately appreciated. Interaction with Joel could be described as nothing short of coronal. Others providing inspiration, ideas, or solid motivation include Fred Hoffman, Yoram Shoham, Leo Buonasera, Jim Robinson, Phil King, Pete Balm, Greg Hindman, and Ken Butler. W. S. "Bill" Killingsworth deserves special credit for providing the final motivation required to chase the dream.

Finally I thank my family. My parents inspired me with an appreciation for academic pursuits long ago and, while I was at UAF, a school was named in honor of my father. My children, Mackenzie and Alexander, have adoringly tolerated a Dad that hid in the back-office way too much, while Marianne was truly the ideal spouse. I could never ask for more, and I dedicate this thesis to them. This one's for you!

Chapter 1

Introduction to the Dissertation

OBJECTIVE

Geotomography is an inverse problem, and is solved using inverse theory. The goal of inverse theory is to estimate optimal model parameters given data and a hypothetical model. The three primary classes of inverse problems are: 1) Over-determined; 2) Under-determined; and 3) Mixed-determined. Unique solutions only exist for over-determined problems. Under-determined and mixed-determined inverse problems require the imposition of *a priori* information for a stable solution to exist.

Geotomography usually requires solving a mixed-determined inverse problem.

The objective of tomography is to image the interior of a body by inverting line-integrals passing through the body for media parameters. Tomography has revolutionized biomedical imaging and is occasionally applied in the field of geophysics (geotomography). It is well known that biomedical tomography produces higher resolution images and more accurate body parameter estimates than seismic geotomography.

Certainly models representing the actual geology of an area are not inconsistent with the data. The apparent limitations of geotomography, therefore, must arise as a consequence of either inadequate solution methods, unsatisfactory models, or some combination thereof.

The investigations reported in this thesis were primarily undertaken to:

- 1) Investigate the fundamental limitations of seismic transmission geotomography;
- 2) Suggest and implement theoretical and computational improvements with application to the advancement of geotomographic imaging methods.

The theory and procedures developed here are demonstrated using both model and field data.

INTRODUCTION TO THE CHAPTERS

Chapter 2 demonstrates that the primary limitation to accurate geotomographic anomaly reconstruction is a direct ramification of the type of *a priori* information used to stabilize the inverse problem. Composite Distribution Inversion is introduced as a means of stabilizing geotomographic systems which still allows estimation of large amplitude, spatially small anomalies. The purpose of Chapter 2 is to introduce CDI as a theory that might extend the range of potential applications appropriate for geotomography. Model data, primarily, are used to illustrate improved imaging potential.

The purpose of **Chapter 3** is to demonstrate CDI tomography applied to the problem of tunnel detection. The importance of accurate source and receiver location information is established. A means of deriving empirical borehole corrections using crosshole data is introduced. The corrected location estimates are used to improve tomographic images of a tunnel near the northern border of South Korea. The tunnel is imaged using both attenuation and traveltime tomography.

Accurate hypocenter estimation and error analysis are required if natural sources are to be reliably used for tomographic imaging. Accurate hypocenter estimation within the context of a three-dimensional velocity model of arbitrary complexity requires rapid, accurate and robust raytracing. The objective of **Chapter 4** is to describe a very robust and accurate method of raytracing based on approximate raypath recovery from a pre-computed traveltimes field followed by refinement to required accuracy. Graph-theory is used for traveltimes field computation. The most efficient method of graph-theoretic queuing described in public domain literature at this time is introduced. Pseudo-bending is optimized and modified for use in general models.

The primary objective of **Chapter 5** is to illustrate efficient and robust earthquake location estimation using data from a limited aperture array, and in the presence of rugged topography. The approach is made practical by using the raytracing method described in Chapter 4. Data from an energetic earthquake swarm at Akutan Volcano, Alaska, were selected for the demonstration data set. The swarm decay sequence source area is found to be spatially small and stationary with respect to time. Additional information documenting the swarm and supporting the interpretation is included.

The purpose of **Chapter 6** is to demonstrate CDI tomography as a potential means of imaging volcanic systems. Concepts from each of the previous chapters are used to produce tomographic velocity estimates at Mt. Spurr, Redoubt, and Augustine Volcanoes, Alaska. The CDI tomograms show that the seismic arrival data residual at these volcanoes could be primarily attributable to a relatively heterogeneous near-surface

features. The data do not require a highly heterogeneous model that extends more than two kilometers below sea-level. No indication of shallow magma storage is found.

SIGNIFICANCE OF THE WORK

The primary contribution of this thesis is the development and demonstration of Composite Distribution Inversion applied to geotomographic imaging. It was previously believed that crosshole geotomography could do no better than 35-50% of point anomaly reconstruction using typical crosshole data (Humphreys and Clayton , 1988). Attenuation is more severe for 3D (Lees and Crosson, 1989). Composite Distribution Inversion (Chapter 2) permits point-anomaly detection and changes the scope of potential geo-imaging problems that may be addressed tomographically.

Tunnel detection has been a geophysical imaging challenge for over 50 years. Data collected during the 1980's and early 1990's, through a research project sponsored by the U.S. Army Corps of Engineers and Department of Defense produced a null result. Data from that project are used here, in conjunction with CDI, to demonstrate the possibility of tunnel detection using seismic crosshole tomography (Chapter 3). A new method for estimating borehole deviation corrections and a new approach to attenuation tomography are also introduced in Chapter 3.

Raytracing is required for a large number of geophysical tasks. The computational demands of raytracing limit the ability to test geophysical hypotheses and the types of models that can be used. Nakanishi and Yamaguchi (1986) suggested a robust algorithm for seismic traveltime field calculation based on graph-theory, but the proposed

implementation was too inefficient for practical application. The purpose of Moser's dissertation (1992) was to improve the efficiency of graph-theoretic traveltime field calculation. The work was a success, and his dissertation thoroughly discusses an efficient method of calculating robust graph-theoretic traveltime fields using a heap-based queue. Later, Moser applied his method to the problem of earthquake location estimation in collaboration with Nolet and van Eck (Moser et al., 1992). Chapter 4 extends and refines Moser's ideas. Though no proof is given, I believe that the graph-theoretic queuing method introduced in Chapter 4 is the most efficient possible method. Dequeueing costs are of order $N \log_2(N)$ for the most efficient methods prior to this work while the method introduced here only incurs a dequeueing cost of order N . Pseudo-bending extensions and optimization are also developed and presented as part of an integrated raytracing methodology.

Chapter 5 uses the raytracing methods developed in Chapter 4 to address the problem of robust earthquake location estimation using typical volcano hazard monitoring arrays. The challenges presented by rugged topography, small arrays and heterogeneous earth models are effectively addressed. The data used to demonstrate the method are from an energetic earthquake swarm and extended seismic decay sequence that occurred during the Spring of 1996 at Akutan Volcano, Alaska. The swarm is significant both because of the cultural hazards posed by Akutan Volcano and because of its extreme intensity relative to other documented swarms. More events occurred during the Akutan swarm than any other swarm in the Global Volcano Earthquake Database (Benoit and McNutt, 1996). The energy released exceeded the suggested eruption warning level of

Yokoyama(1988) by seven orders of magnitude. The study demonstrates a number of reusable concepts that might be applied during future assessment of volcano-related hazards during an ongoing seismic decay sequence.

Chapter 6 is an operational application of the methods developed in the previous chapters. The imaging objectives are three Alaskan volcanoes. The models developed are non-unique alternatives to estimates constructed using smoothness or damping.

All software used was developed as part of an abstract, C++ tomography research library. The code volume now includes over 1000 files and 100,000 lines of code. The software is ANSI C++ and is currently running on a variety of platforms. It is a unique platform for future seismic tomography research.

CONCLUSIONS

The primary conclusions drawn from this work are:

- 1) CDI tomography can be used to detect small and discrete anomalies;
- 2) Crosshole tunnel detection using CDI tomography is possible;
- 3) Tomographic image resolution is strongly dependent upon the accuracy of source and receiver location information;
- 4) Accurate, robust, and efficient raytracing can be implemented by refining raypaths derived from approximate traveltimes fields;
- 5) Nonlinear optimization is required to estimate optimal hypocenter locations using general 3D velocity models and efficient, accurate raytracing is required to make such an approach practical;

- 6) The methods described above can be used in conjunction with both natural and explosion sources to estimate the velocity structure of volcanoes.

RECOMMENDATIONS FOR FUTURE INVESTIGATIONS

CDI encourages small anomalies to approach a state of zero perturbation and allows large anomalies to grow to whatever amplitude is required to effectively reduce prediction error. The assumptions made in the theoretical development of CDI are similar to the ideas used in the statistical field of mixture models. Further integration of CDI within the context of mixture models is appropriate.

CDI has only been applied to tomographic anomaly detection. The theory is applicable to other inverse problems. Non-tomographic mixed-determined inverse problems concerned with anomaly detection might consider use of CDI.

The raytracing methods developed in this thesis are very efficient and robust, but have only been applied to earthquake location and tomography. Mainstream geophysical methods, such as depth migration, could benefit from the raytracing methodology presented here.

The tools required to accurately image the detailed structure of an active volcanic cone have been developed and demonstrated. A large multichannel data acquisition experiment could provide data that would allow detailed CDI tomographic imaging of a volcanic cone. No detailed image of any active volcanic cone has been produced to date. Appropriate data and CDI tomography could resolve hypothesized ideas about the shallow structure and mechanics of volcanoes.

MATERIAL CONTRIBUTORS TO THIS WORK

Doug Christensen was involved with all aspects of this work as the thesis advisor. Richard Rechtien provided the data used to demonstrate tunnel detection. The U.S. Army Corps of Engineers granted permission to publish the tunnel data. John Lahr provided phase data, format information, and assistance with earthquake location concepts. Phil Dawson provided data from the Redoubt temporary array. Juergen Kienle (primarily) acquired the 1977 Augustine shot data, and provided it for use in this investigation. The University of Alaska Fairbanks Geophysical Institute provided student and intern volunteers to conduct the 1995 Augustine shot experiment. Hans Pulpan, Doug Lalla, and John Power provided historical information about Augustine seismicity. John Power provided data from the main Akutan swarm, his picks of coda duration (which provided a much needed cross-check), and digitized contours used in the volcano figures. The Alaska Volcano Observatory (AVO) provided archived phase data and computational facilities. John Eichelberger and Terry Keith provided office space and the opportunity to interact with AVO staff.

Projects were funded in part by NSF grants EAR91-18090, EAR-94-05471, EAR-95-06379, the Alaska Volcano Observatory, and United States Geological Survey grant 14-08-0001-A0574 administered through the Alaska Volcano Observatory. Scholarships were provided by the Alaska Geophysical Society and Exxon through the SEG Scholarship Foundation.

REFERENCES

- Benoit, J. P., and McNutt, S. R., 1996, Global volcanic earthquake swarm database and preliminary analysis of volcanic earthquake swarm duration: *Annali de Geofisica*, **39**, 221-229.
- Humphreys, E., and Clayton, R. W., 1988, Adaptation of back projection tomography to seismic travel time problems: *J. Geophys. Res.*, **93**, 1073-1085.
- Lees, J. M., and Crosson, R. S., 1989, Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data: *J. Geophys. Res.*, **94**, 5716-5728.
- Moser, T. J., 1992, The shortest path method for seismic ray tracing in complicated media, Ph.D. thesis, Institute of Geophysics, Utrecht University, Netherlands.
- Moser, T. J., van Eck, T., and Nolet, G., 1992, Hypocenter determination in strongly heterogeneous earth models using the shortest path method: *J. Geophys. Res.*, **97**, 6563-6572.
- Nakanishi, I., and Yamaguchi, K., 1986, A numerical experiment on nonlinear image reconstruction from first-arrival times for two-dimensional island arc structure: *J. Phys. Earth*, **34**, 195-201.
- Yokoyama, I., 1988, Seismic energy release from volcanoes: *Bull. Volcan.*, **50**, 1-13.

Chapter 2

Composite Distribution Inversion Applied to Crosshole Tomography

ABSTRACT

Crosshole tomography requires solution of a mixed-determined inverse problem and addition of *a priori* information in the form of auxiliary constraints to achieve a stable solution. Composite distribution inversion (CDI) constraints are developed by assuming parameters are drawn from a composite distribution consisting of both normally and uniformly distributed parameters. Non-anomalous parameter estimates are assumed to be Gaussian while anomalous parameters are assumed uniform. The resulting constraints are sensitive to anomaly volume and are an alternative to the usual constraints of minimizing L_2 solution length or some measure of roughness. Damped least-squares inversion, which minimizes solution length, distributes anomalous signal through poorly resolved areas to produce attenuated and smoothed anomalies. Similar regularization methods, such as smoothness or flatness constraints, also degrade small spatial wavelength features and produce diffuse images of distinct anomalies. CDI constraints preserve small spatial wavelength features by encouraging small amplitude anomalies to assume the value of the reference model, and by allowing truly anomalous parameter estimates to assume whatever value minimizes prediction error without incurring additional penalty. CDI tomograms are characterized by nearly ideal point-spread functions, suggesting the possibility of better quantitative parameter estimates than are produced using most existing methods. CDI tomograms of both synthetic and field data

are shown to produce less diffuse images with more accurate anomaly amplitude estimates than damped least-squares methods. The CDI algorithm is potentially applicable to non-tomographic inversion problem.

INTRODUCTION

Tomographic imaging (Gilbert, 1972; Hounsfield, 1973) has dramatically improved medical diagnostics and is one of the more significant geophysical technology improvements of the 1980's (Lawyer, 1992). While modern medical tomograms are high quality, the resolution and quality of subsurface images are generally poor by comparison (Singh and Singh, 1991). The degraded images of geotomography are a result of problem formulation, physical limitations of acquisition geometry, and constraints imposed to achieve stable solutions.

Conventional tomography requires solving the system $\begin{bmatrix} \mathbf{A} \\ \mathbf{W} \end{bmatrix} \mathbf{s} = \begin{bmatrix} \mathbf{t} \\ \mathbf{c} \end{bmatrix}$, where \mathbf{A} is a computed pathlength matrix, \mathbf{t} is a data vector (e.g. traveltimes or amplitudes), \mathbf{s} is a parameter vector to be estimated (e.g. velocity or attenuation exponent), and $\mathbf{W} \mathbf{s} = \mathbf{c}$ represents a system of auxiliary constraints imposed to achieve a stable solution. In this form there are three commonly made assumptions which limit image quality:

- (1) Sufficient parameterization;
- (2) Validity of ray theory and raytracing approximations;
- (3) Limited image degradation resulting from the imposition of auxiliary constraints.

Large discretization errors and/or ray-tracing inaccuracies may indicate an insufficient parameterization. These problems may be addressed by using smaller cells, or by using a continuous parameterization with ray bending (Thurber and Ellsworth, 1980; Um and Thurber, 1987). The assumption that rays represent the travel path followed by first arrival energy implies infinite frequency wave propagation (Wielandt, 1987). When this is not valid, diffraction tomography is more appropriate. The final assumption, limited image degradation as a consequence of auxiliary constraints, is usually satisfied when imaging smooth structures, however, images of small spatial wavelength features are often quite degraded. Small, distinct, anomaly amplitudes are severely attenuated, and anomalous signal is distributed through the model, resulting in streaking. This type of image degradation is a direct consequence of assumptions made to stabilize the inverse problem.

Most commonly, the inverse is stabilized by introducing an additive damping value. Other stabilizing constraints may be used as an alternative to, or in conjunction with, simple damping. Commonly used methods include use of minimum roughness constraints (Shaw and Orcutt, 1985; Scales et al., 1990), spatial averaging (Evans and Zucca, 1988), convolutional quelling (Meyerholtz et al., 1989), and spatial averaging with outlier elimination (Gersztenkorn and Scales, 1988). These methods improve some aspects of image quality by decreasing undesirable parameter value fluctuations but degrade other aspects by attenuating anomaly amplitudes and smoothing anomaly boundaries. It is recognized that smooth estimates are not always desirable and some recent work has been focused on producing sharper, rather than smoother, images.

'Tightening' constraints guided by interpretive assumptions about anomaly location are shown to produce sharp images of distinct features in a known background (Carrion, 1991). Singh and Singh (1991) also produce sharp images of distinct features in a known background using interpretive constraints guided by assumptions of anomaly polarity. Spatial-coherency filtering is shown to preserve spatially coherent small wavelength features (Zhou, 1993). The combination of hard deterministic constraints and soft Gibbs constraints is promising, and synthetic model studies demonstrate superior imaging of isolated anomalies (Carrion et al., 1993).

The objective of this paper is to present a set of auxiliary constraints which allow reliable detection of small spatial wavelength features while maintaining a stable inverse. We first outline basic problem formulation and then develop auxiliary constraints by assuming *a priori* knowledge of non-anomalous parameter values and complete ignorance of anomalous values. This results in our composite distribution inversion (CDI) strategy. We then describe algorithm implementation tactics, followed by examples of application to both synthetic and field data.

METHODS

Problem formulation

One approach to geotomography begins with the basic equation:

$$t_i = \sum_{j=1}^N A_{ij} S_j \quad \text{for } i = 1, N \quad (1)$$

where: t_i = datum associated with the ray i (traveltime or amplitude)

A_{ij} = the length of ray i in cell j

s_j = parameter value of cell j (slowness or absorption)

or in matrix form,

$$\underline{\mathbf{t}} = \underline{\mathbf{A}} \underline{\mathbf{s}}. \quad (2)$$

When equation (2) is solved directly, the squared solution length is given by $\underline{\mathbf{s}}^T \underline{\mathbf{s}}$, and constraints involving solution length seek a model with minimum overall parameter estimates. This is not desirable, and provides motivation for reformulating the problem in terms of a background model and perturbation values, exactly as is done when linearizing a nonlinear problem but for quite different reasons. We will use a measure of solution length rather than flatness or roughness (which may be appropriately measured by parameter estimates directly), and therefore choose to use a reference model and solve for perturbations. The basic equation we solve is

$$\Delta \underline{\mathbf{t}} = \underline{\mathbf{A}} \Delta \underline{\mathbf{s}}, \quad (3)$$

where $\Delta \underline{\mathbf{s}} = \underline{\mathbf{s}} - \underline{\mathbf{s}}_0$, $\Delta \underline{\mathbf{t}} = \underline{\mathbf{t}} - \underline{\mathbf{t}}_0$, and $\underline{\mathbf{t}}_0$ is a vector of theoretical data values associated with the reference model, $\underline{\mathbf{s}}_0$. Any of several methods might be used to solve equation (3) for $\Delta \underline{\mathbf{s}}$, including the direct or singular value decomposition (SVD) computed generalized inverse (Golub and Reinsch, 1970; Lines and Treitel, 1984; Lines and LaFehr, 1989), back projection methods such as ART (Kaczmarz, 1937; Tanabe, 1971; Dines and Lytle, 1979) or SIRT (Gilbert, 1972; Ivansson, 1986), and sparse system conjugate gradient methods (Hestenes and Stiefel, 1952; Paige and Saunders, 1982;

Scales, 1987). For most tomography problems, \mathbf{A} is ill-conditioned, and direct solution results in unstable model parameter estimates. To stabilize the solution, *a priori* information must be added. This information is usually specified in the form of auxiliary constraints implemented via weighting matrices. The objective function to be minimized is given by:

$$E = \left(\mathbf{A} \Delta \mathbf{s} - \Delta \mathbf{t} \right)^T \mathbf{W}_e \left(\mathbf{A} \Delta \mathbf{s} - \Delta \mathbf{t} \right) + \Delta \mathbf{s}^T \mathbf{W}_m \Delta \mathbf{s}, \quad (4)$$

where \mathbf{W}_m and \mathbf{W}_e are model and error weighting matrices, respectively.

If we define $\mathbf{W}_m \equiv \mathbf{D}_m^T \mathbf{D}_m$ and $\mathbf{W}_e \equiv \mathbf{D}_e^T \mathbf{D}_e$, then an L_2 penalty function interpretation of E is

$$E = \left\| \begin{bmatrix} \mathbf{D}_e & \mathbf{A} \\ \mathbf{D}_m & \mathbf{I} \end{bmatrix} \Delta \mathbf{s} - \begin{bmatrix} \mathbf{D}_e \Delta \mathbf{t} \\ \mathbf{0} \end{bmatrix} \right\|_2. \quad (5)$$

For diagonal \mathbf{D}_e , weighted least-squares parameter estimates may be written as

$$\Delta \mathbf{s} = \left(\mathbf{A}^T \mathbf{D}_e^T \mathbf{D}_e \mathbf{A} + \mathbf{D}_m^T \mathbf{D}_m \right)^{-1} \mathbf{A}^T \mathbf{D}_e^T \mathbf{D}_e \Delta \mathbf{t}. \quad (6)$$

When \mathbf{D}_m and \mathbf{D}_e are not functions of $\Delta \mathbf{s}$ or $\Delta \mathbf{t}$, respectively, equation (6) may be solved with one matrix inversion, or alternatively one pass through an iterative least-squares equation solver. If \mathbf{D}_m or \mathbf{D}_e are functions of $\Delta \mathbf{s}$ or $\Delta \mathbf{t}$, an iteratively reweighted least-squares approach may be used (Byrd and Pyne, 1979), and non- L_2 norms are possible. These methods allow considerable flexibility in specifying *a priori*

information. The primary task, given methods to solve the preceding equations, is to select *a priori* information with physically desirable ramifications.

Appropriate *a priori* information

In this section we briefly review standard *a priori* assumptions and ramifications. We then modify the usual assumptions by assuming parameters are drawn from a composite distribution consisting of two sub-populations. One sub-population is associated with non-anomalous values and is distributed normally. The other is associated with anomalous values and distributed uniformly. The resulting constraints encourage non-anomalous parameter estimates while still maintaining minimal prediction error.

When data are assumed Gaussian (normally distributed), $\mathbf{W}_e = [\text{cov } \Delta \mathbf{t}]^{-1}$ (Menke, 1984). This is a good assumption if a variety of processes contribute errors of similar magnitude to the total error because the total distribution approaches Gaussian as the number of processes becomes large, regardless of the individual distributions (Hogg and Craig, 1978). If model parameters are assumed *a priori* to be Gaussian then,

$\mathbf{W}_m = [\text{cov } \Delta \mathbf{s}]^{-1}$. If all covariances are zero, then the weighting matrices are diagonal with data and model variances on the respective main diagonals.. If it is further assumed that data are of equal variance, σ_d^2 , and that model parameters are of equal variance, σ_m^2 , then the usual damped least-squares solution is appropriate.

In some cases parameter estimates may be improved by changing one or more of the assumptions leading to the least-squares solution. Noisy outliers, for example, suggest a

single large amplitude noise source which may not be Gaussian. When noisy outliers are present, the assumption of exponentially distributed data (which implies use of the L_1 norm) results in less noise-sensitive parameter estimates (Claerbout and Muir, 1973). Other commonly made assumptions which deviate from standard least-squares include specification of non-zero model parameter covariances in such a way that the solution is constrained to be smooth or flat.

We propose that the assumption of normally distributed parameters may be improved when attempting to image anomalies in a relatively well known background. We assume that parameters are drawn from a composite distribution. Parameters with a value near the background reference are normally distributed as $n(\mu, \sigma)$, that is normal with a mean of μ and standard deviation of σ . If perturbations are used, as in equation (3), the parameter perturbations associated with background values are assumed to be distributed as $n(0, s)$. For these parameters, $\mathbf{W}_m = [\text{cov } \Delta \mathbf{s}]^{-1}$, and the appropriate weights are given by $D_{mii} = \sigma^{-1}$. Now consider a state of total ignorance concerning anomalous parameter values. This may be described by a uniform distribution with probability density equal to some constant over an interval. In this case all parameter values are equally likely and the second term in equation (4), $\Delta \mathbf{s}^T \mathbf{W}_m \Delta \mathbf{s}$ (a weighted measure of solution length), should be constant. This is satisfied when $\mathbf{D}_m = \text{diag}(\alpha \Delta s_i^{-1})$, where α may be chosen arbitrarily. Now assume that some fraction, p_1 , of the parameters to be estimated belong to the assumed Gaussian non-anomalous distribution and that a state of complete

ignorance applies to the remaining fraction, $p_2 = 1 - p_1$. Given this composite distribution, the expected value for \mathbf{D}_m is given by

$$\mathbf{D}_{m_{ii}} = \frac{p_1 f_1(\Delta s_i) \sigma^{-1} + p_2 \alpha \Delta s_i^{-1} f_2}{p_1 f_1 + p_2 f_2} \quad (7)$$

where: $f_1(\Delta s_i)$ = probability density function for the distribution $n(\mu, \sigma)$

$$= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(\Delta s_i - \mu)^2}{2\sigma^2}\right)$$

f_2 = probability density for a uniform distribution, constant.

Equation (7) is a model parameter weighting function for use when estimating parameters from a composite distribution as described above, though a minor modification is required for computational use. The modification is required because computational singularities are encountered for small Δs_i , which may be addressed by replacing Δs_i with $\Delta s_i + \epsilon$, where ϵ is small. A similar problem is encountered when using an iteratively reweighted approach to achieve an L_1 norm. Scales et al.(1988) implement the L_1 norm by letting $\mathbf{W}_e = \text{diag}(\Delta r_i^{-1})$, or equivalently $\mathbf{D}_e = \text{diag}(\alpha \Delta r_i)$, where Δr_i is the prediction error residual associated with the i th datum. They replace Δs_i with ϵ when $\Delta s_i < \epsilon$, which is also suggested by Huber (1981). For our application, we prefer replacing Δs_i with $\Delta s_i + \epsilon$. Either approach produces approximately the same result.

The last modification to equation (7) is to specify f_2 in terms of f_1 to facilitate ease of use in parameter specification. The idea is to allow the user to ensure an appropriate relative value of f_2 relative to f_1 by specifying a density function crossover at some point on the cumulative distribution function. Consider the two-sided normal cumulative distribution,

$$\begin{aligned} F_1(\Delta s) &= \int_{-\Delta s}^{\Delta s} f_1(x) dx \\ &= \text{erf}\left(\frac{|\Delta s - \mu|}{\sqrt{2} \sigma}\right), \end{aligned} \quad (8)$$

with an inverse such that

$$\Delta s = F_1^{-1}(F_1(\Delta s)), \quad (9)$$

where

$$F^{-1}(p) = \sqrt{2} \sigma \text{erf}^{-1}(p) \pm \mu \quad (10)$$

and

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-r^2} dr. \quad (11)$$

This simply specifies that the cumulative distribution probability value, c , may be inverted for Δs such that $c = F_1(\Delta s)$. Using these relationships, the final weighting function is given by

$$D_{mii} = \frac{\frac{p_1 f_1(\Delta s_i)}{\sigma} + \frac{p_2 f_2 \alpha}{[\Delta s_i + \varepsilon \text{sgn}(\Delta s_i)]}}{p_1 f_1 + p_2 f_2} \quad (12)$$

with

$$f_2 = f_1(F_1^{-1}(c)) = \text{constant} \quad (13)$$

Characteristics of the penalty function implied by equation (12) for small and moderate perturbations are illustrated for one specific choice of parameters in Figure 1a. No penalty is assessed for zero-valued parameter perturbation estimates. In the zone of small, but non-zero values, the penalty rises more sharply than a Gaussian penalty function. This results in a strong tendency for small perturbations to assume a true zero value when compared with damped least-squares. The next region of the curve, in the area of moderate parameter perturbations, is almost linear. This would be expected if exponentially distributed parameters were assumed and indicates a distribution with thicker tails than a normal curve. For larger perturbations (Figure 1b), the penalty value asymptotically approaches a constant, specifying little preference for one value as opposed to another. This contrasts sharply with the Gaussian penalty function, which is characterized everywhere by a linearly increasing preference for smaller values as perturbation value increases, that is the derivative of the penalty function with respect to perturbation magnitude is constant for all perturbation values.

Conceptually, CDI constraints bear some resemblance to principles behind minimum entropy deconvolution, commonly referred to as MED (Wiggins, 1978; Ulrych and Walker, 1982; Cabrelli, 1984). MED seeks a simple solution, characterized by a small number of anomalous values, implemented via the varimax norm of factor analysis. The basic idea of MED is to discourage parameter estimates that differ appreciably from zero. CDI encourages small anomalies to assume a nil state, while more or less allowing larger anomalous estimates to float in response.

Implementation

Since \mathbf{D}_m as defined in equation (12) is a function of the parameter estimates, a non-linear inverse procedure is required. A simple method to implement nonlinear inversion is to use an iteratively reweighted application of linear least-squares. Parameter estimates at reweight iteration k are given by

$$\Delta \mathbf{s}^{(k)} = \left(\mathbf{A}^T \mathbf{W}_e^{(k)} \mathbf{A} + \mathbf{W}_m^{(k)} \right)^{-1} \mathbf{A}^T \mathbf{W}_e^{(k)} \Delta \mathbf{t} \quad (14)$$

with

$$D_{m_{ii}}^{(k)} = \frac{\frac{p_1 f_1(\Delta s_i^{(k)})}{\sigma} + \frac{p_2 f_2 \alpha}{\left[\Delta s_i^{(k)} + \varepsilon \operatorname{sgn}(\Delta s_i^{(k)}) \right]}}{p_1 f_1 + p_2 f_2} \quad (15)$$

A slightly overdamped least-square solution is chosen as the first parameter estimate, $\Delta \mathbf{s}^{(1)}$. Given the initial parameter estimate, constraint weights are computed using equation (15). Revised parameter estimates are then computed by solving equation (14) for $\Delta \mathbf{s}^{(2)}$. This process is repeated until there is little change in parameter estimates between iterations, at which point convergence is assumed. This procedure is identical to that used to implement the general L_p norm solution (Scales et al., 1988). A formal proof for convergence for this type of problem is given by Eckhart (1980). Note that iterative application of the diagonal weights specified in equation (15) may be used in conjunction with other weighting functions, including L_p norm weights, noise weighting, or with off-diagonal weights if smoothing is desired.

RESULTS

Evaluation methods

Our objective is to assess the potential performance of composite distribution inversion as applied to crosshole tomography. Established diagnostic velocity distributions are used as well as one application to a field data set. The model studies use fewer rays than unknowns to ensure a large underdetermined component for test purposes. We seek to not only demonstrate unequivocal successes of CDI tomography, but also to illustrate limitations of the method. Our hypothetical distributions include 10% velocity anomalies and traveltimes data with straight rays. While not strictly valid, the straight ray approximation is regarded as acceptable for velocity contrasts $< 15\%$ (Singh and Nyland, 1988). Motivation for using straight rays is provided by considering that our goal is to compare intrinsic imaging capabilities given fixed pathlength matrices. Use of kinked rays would obfuscate objective comparison of constraint effectiveness by producing results which include both the effects of different auxiliary constraints, as well as different pathlength matrices.

In the following synthetic examples we have set $p_1 = p_2 = 1/2$. Total prediction error for all cases compared is nearly equal, indicating variations in the solution consist primarily of null-space vectors. After testing the method using synthetic data, we apply CDI tomography to a field data set acquired for tunnel detection purposes. The area covered by this data set is small, and straight rays are assumed to be adequate for evaluation purposes.

Point anomalies

Some fundamental limitations of imaging algorithms may be evaluated using point-spread functions (Humphreys and Clayton, 1988). The point-spread function for any given location within the imaging area (or volume) is defined as the tomogram which results from inverting theoretically calculated data associated with a point anomaly at the location. Point-spread functions are spatially variable (due to variable raypath coverage) and provide diagnostic information similar to that provided by impulse response functions used in other types of analysis.

Point anomalies are the simplest possible features of a tomogram. An ideal inverse should be characterized by an ideal point-spread function regardless of ray inhomogeneity or uneven angular coverage. Synthetic traveltime data generated using the model in Figure 3a and ray geometry of Figure 2 are used to evaluate this performance objective, with the resulting CDI tomogram shown in Figure 3c. The model is reproduced to within stopping criteria precision. Neither point anomaly is significantly distorted or degraded in amplitude, though both are influenced by varying amounts of uneven angular coverage. This performance contrasts sharply with the damped least-square tomogram result (Figure 3b), which is characterized by anomaly attenuation and distortion.

Anomalous model bodies

High contrast 'brick' anomalies were previously used to study the performance of standard tomography algorithms (Philips and Fehler, 1991) as well as innovative constraint methods (Singh and Singh, 1991; Carrion, 1991; Carrion et al., 1993; Carrion et al., 1997). We use the model shown in Figure 4a, which is similar to that used by Philips and Fehler (1991), to test the performance of CDI tomography in the presence of high contrast distinct anomalies. The model is an area of 25 blocks by 25 blocks, with each block having dimensions 40m x 40m. The velocity field is described by

$$V = \begin{cases} V_2, & \text{inside square with side } c \\ V_1, & \text{otherwise,} \end{cases}$$

where the anomalous area is centered and has sides, $c = 200$ m, $V_1 = 4000$ m/s and $V_2 = 3600$ m/s. Sampling raypaths were generated by placing 20 equally spaced sources on one side of the imaging area, and 20 equally spaced receivers on the other. Each source is connected to each receiver by a raypath (a total of 400 rays). This set of rays is used to image the 625 cells. A damped least-square tomogram for this test case is shown in Figure 4b. The general anomalous zone is easily identifiable, though areas of significant distortion are obvious. The CDI tomogram is shown Figure 4c. The model is reproduced to within numerical stopping criteria, and offers significant improvement over damped least-squares.

Both the point and brick anomalies discussed above are isolated constant velocity anomalies, and excellent performance of CDI is expected. Gradational anomalies present a more challenging test because truly anomalous cells with a small perturbation value could be incorrectly pushed to a background value. Figure 5a is a hypothetical 'blob'

model, also similar to that used by Philips and Fehler (1991). The velocity distribution is described by:

$$V = \begin{cases} V_2 + (V_2 - V_1)(r / r_0), & r < r_0 \\ V_1, & r \geq r_0, \end{cases}$$

with $V_1 = 4000$ m/s, $V_2 = 3600$ m/s, r = distance from the center of the model in cells, and $r_0 = 8$ cells. The ray geometry is the same as that employed for the brick anomaly described in the previous section. Characteristic damped least-square distortion is observed (Figure 5b). The magnitude of the distortion is somewhat less on the CDI tomogram (Figure 5c), which also retains more accurate shape and magnitude information. Tests performed with twice the number of rays used above (800 rays, not shown) resulted in improved, but still quite distorted, least-squares images, and nearly perfect CDI images. As the underdetermined components of a problem increases, both Gaussian and CDI methods break down (in terms of producing a desirable image), though CDI is significantly more robust.

As a final model test, we consider a distinct anomaly with concave boundaries, both with and without noise (Figure 6a). The anomaly was sampled by the raypaths in Figure 1. Typical distortion and attenuation are present on the noise-free damped least-squares tomogram (Figure 6b). The noise-free CDI tomogram reproduces the original model (Figure 6c). Zero-mean Gaussian noise was then added, with a standard deviation equal to one-fourth the traveltimes anomaly associated with one anomalous cell. The slightly underdamped least-square tomogram is quite noisy (Figure 6d), while the critically damped tomogram is smoother but highly attenuated (Figure 6e). The CDI tomogram

(Figure 6f) retains better amplitude information and suppresses much of the low amplitude noise, although the image is still quite distorted. This demonstrates that while CDI is useful primarily as a tool to reduce non-uniqueness without smoothing outliers, it is somewhat effective at suppressing low amplitude noise. CDI is neither designed nor appropriate for use as a high amplitude noise reduction tool. Use of noise weighting and/or an L_1 prediction error norm must be used in conjunction with CDI in the presence of high noise levels.

Field data application

Seismic data collected in two boreholes for the purpose of tunnel detection are used to investigate the performance of CDI tomography when applied to field data with inherent noise. Lithology at the site was uniform granite. A 2 m diameter tunnel was known to exist between the boreholes. Data were collected using a series of seven data acquisition runs, each with a fixed source-receiver vertical offset. A run was started by lowering both the source and receiver to initial positions in their respective boreholes, and collecting traveltimes between these two points. Both the source and receiver were then raised 0.2m to a new data acquisition position. During each run, the process of raising both the source and receiver at 0.2m increments and observing traveltimes was repeated until data were collected at 150 source-receiver locations. The seven acquisition runs resulted in a total of 1050 raypaths collected at seven source-receiver azimuths (Figure 7). The source was a high frequency arc-discharge type with a bandwidth of 300 - 7000 Hz and a dominant frequency of approximately 1500 Hz (Rechtien et al., 1993).

The area to be imaged was divided into 40 vertical cells and 17 horizontal cells, with each cell having dimensions of (1.164 m x 1.164 m). Mean velocity was used as the background model (5386 m/s).

Two nearly critically damped least-squares tomograms are presented, one slightly underdamped, one perhaps slightly overdamped, subject to interpretation. In both cases there is a diffuse low velocity zone near the center as well as several noise related artifacts (Figure 8a, b). Features aligned with acquisition azimuths are interpreted as noise related features. The slightly overdamped case attenuates some noise related features, but also attenuates and smears the target anomaly (Figure 8a). The slightly underdamped target is less attenuated, but noise artifacts begin to assume relatively high amplitudes (Figure 8b). The CDI tomogram (Figure 8c) is characterized by better target amplitude estimates and much less noise in the background. Eleven reweight iterations were required for convergence, with most of the noise suppression occurring in the first five iterations (Figure 8d). RMS traveltimes errors of each solution are all approximately nil: CDI rms error = 1.1×10^{-7} s; slightly overdamped = 1.1×10^{-7} s; slightly underdamped = (9.9×10^{-8}) s).

The differences between the field tomograms may be further elucidated by examining the distribution of estimated velocities (Figure 9). A connected line histogram of estimated values shows the overdamped tomogram is distributed as a unimodal function about the mean with small tails. The underdamped tomogram is also unimodal, but with longer, thicker tails associated with the noisy estimates. The CDI histogram is bi-modal,

but is more narrowly peaked at the primary mode. These phenomena arise because damped least-squares methods assign larger penalties as parameter estimates move away from the mean, thus encouraging a single mode. In contrast, CDI essentially pushes small amplitude anomalies to the mean, while allowing truly anomalous parameters to assume whatever value is required to minimize prediction error without increased penalty. This is consistent with the original premises of our development, specifically that *a priori* estimates of non-anomalous values is reasonable, but a state of total ignorance concerning anomalous parameter values may often apply

DISCUSSION AND CONCLUSIONS

Underdetermined components in geotomographic problems force the imposition of *a priori* information to achieve stable solutions. CDI provides a relatively general alternative to other commonly used *a priori* information, which does not preclude imaging small spatial wavelengths. Conceptually, CDI is appropriate any time the objective is to image anomalous features with unknown parameter values in a normally distributed background. The principles of CDI are potentially applicable to other inverse problems with similar characteristics.

Point-spread functions are critical diagnostics for tomography algorithm performance assessment, analogous to impulse response functions in signal processing applications. Damped least-square point-spread functions were previously assumed to represent a best case with a performance that other algorithms should attempt to achieve (Humphreys and Clayton, 1988). CDI offers a substantial improvement in the form of ideal point-spread

functions, free of anomaly attenuation and smoothing. Reduced anomaly attenuation may be most beneficial in three-dimensional applications, where acceptable point-spread function amplitudes are as small as 15% of true amplitudes (Lees and Crosson, 1989).

The CDI imaging improvements demonstrated here cannot be replicated by simply adjusting the parameters of damping or smoothness-based regularization methods. Increased damping and smoothness decrease noise and reduce estimation variance primarily by attenuating small wavelength features. Decreased damping reduces attenuation and smoothing of small wavelengths, but fails to maintain sufficient *a priori* information to achieve a stable solution, resulting in increased noise. CDI maintains sufficient *a priori* information to keep the overall image near the background model, though sharp features may be imaged because large penalties are not assessed when a high amplitude feature is needed to satisfy experimental data.

Most previous methods of sharpening tomographic images rely on explicit assumptions about the anomalous bodies present (Carrion, 1991; Singh and Singh, 1991) or produce noise-dominated images which result from simply trading variance for resolution through constraint reduction. CDI contrasts with explicit assumption methods by actually assuming a greater state of ignorance than damped least-squares. CDI tomograms demonstrate that it is possible to introduce sufficient *a priori* information to achieve a reasonable, stable solution, while maintaining accurate amplitude estimates and the power to image small spatial wavelengths. Future research possibilities include investigation of CDI in conjunction with noise weighting, enhancement CDI algorithm efficiency, or application of CDI to appropriate non-tomographic problems.

ACKNOWLEDGEMENTS

Funding for this project was provided in part by United States Geological Survey grant 14-08-0001-A0574 administered through the Alaska Volcano Observatory.

Scholarship support was provided by the Alaska Geophysical Society through the SEG Scholarship Foundation. We extend thanks to Joel Hayden, Keith Echelmeyer, Hans Pulpan, Juergen Kienle, and Ron Barry for constructive review of key concepts, and to Dapeng Zhao for his helpful presence.

We also thank Robert F. Ballard for his assistance in making the field data shown in Figure 8 available. Field data were collected under the Tunnel Detection Program sponsored by the U.S. Army Belvoir Research, Development, and Engineering Center, Fort Belvoir, Virginia, and the U.S. Waterways Experiment Station, Vicksburg, Mississippi. Permission was granted by the Chief of Engineers to publish this data. The views of the authors do not purport to reflect the position of the Department of the Army or the Department of Defense.

REFERENCES

- Byrd, R. A., and Pyne, D. A., 1979, Convergence of the iteratively reweighted least-squares algorithm for robust regression: Technical Rep. No. 313, The Johns Hopkins Univ.
- Cabrelli, C. A., 1984, Minimum entropy deconvolution and simplicity: A noniterative algorithm: *Geophysics*, **50**, 394-413.
- Carrion, P., 1991, Dual tomography for imaging complex structures: *Geophysics*, **56**, 1395-1404.
- Carrion, P., Jacovitti, G., and Neri, A., 1993, Gaussian and non-Gaussian tomographic modelling via simulated annealing: *Journ. Seis. Expl.*, **2**, 189-204.
- Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: *Geophysics*, **38**, 826-844.
- Dines, K. A., and Lytle, R. J., 1979, Computerized geophysical tomography, *Proc. IEEE*, **67**, 1065-1073.
- Eckhart, U., 1980, Weber's problem and Weisdeld's algorithm in general spaces: *Math. Programming*, **18**, 186-196.
- Evans, J. R., and Zucca, J. J., 1988, Active high-resolution seismic tomography of compressional wave velocity and attenuation structure at Medicine Lake volcano, northern California Cascade range: *J. Geophys. Res.*, **93**, 15016-15036.
- Gersztenkorn, A., and Scales, J. A., 1988, Smoothing seismic tomograms with alpha-trimmed means: *Geophys. J.*, **92**, 67-72.

- Gilbert, P., 1972, Iterative methods for the three-dimensional reconstruction of an object from projections: *J. Theor. Biol.*, **36**, 105-117.
- Golub, G., and Reinsch, C., 1970, Singular value decomposition and least-squares solutions: *Numer. Math.* **14**, 403-420.
- Hestenes, M., and Stiefel, E., 1952, Methods of conjugate gradients for solving linear systems: *Nat. Bur. Standards J. Res.*, **49**, 409-436.
- Hogg, R. V., and Craig, A. T., 1978, *Introduction to Mathematical Statistics*, 4th edition: Macmillan Publishing Co., Inc.
- Hounsfield, G. N., 1973, Method and apparatus for measuring X or gamma radiation absorption or transmission at plural angles and analyzing the data: U.S. Patent 3 778 614.
- Huber, P., 1981, *Robust Statistics*: John Wiley and Sons, Inc.
- Humphreys, E., and Clayton, R. W., 1988, Adaptation of back projection tomography to seismic travel time problems: *J. Geophys. Res.*, **93**, 1073-1085.
- Ivansson, S., 1986, Seismic borehole tomography - Theory and computational methods: *Proc. IEEE*, **74**, 328-338.
- Kaczmarz, S., 1937, AngenŠherte Auflšssung von Systemen linearer Gleichungen: *Bull. Acad. Polon. Sci. Lett.*, **A**, 355-357.
- Lawyer, L.C., 1992, Times of change - the last decade: Paradox of the '80s, *The Leading Edge*, **11**, no. 6, 30-32.
- Lees, J. M., and Crosson, R. S., 1989, Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data: *J. Geophys. Res.*, **94**, 5716-5728.
- Lines, L. R., and Treitel, S., 1984, A review of least-squares inversion and its application to geophysical problems: *Geophys. Prosp.*, **32**, 159-186.

- Lines, L. R., and LaFehr, T., 1989, Tomographic modeling of a cross-borehole data set: *Geophysics*, **54**, 1249-1257.
- Menke, W., 1984, *Geophysical data analysis: Discrete inverse theory*: Academic Press Inc.
- Meyerholtz, K., Pavlis, G., and Szpakowski, S., 1989, Convolutional quelling in seismic tomography: *Geophysics*, **54**, 570-580.
- Paige, C. C., and Saunders, M. A., 1982, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.* **8**, 43-71.
- Philips, W. S., and Fehler, M. C., 1991, Traveltime tomography: a comparison of popular methods: *Geophysics*, **56**, 1639-1649.
- Rechtien, R. D., Hambacker, K. L., and Ballard Jr., R. F., 1993, A high frequency sparker source for the borehole environment: *Geophysics*, **58**, 660-669.
- Scales, J. A., 1987, Tomographic inversion via the conjugate gradient method: *Geophysics*, **52**, 179-185.
- Scales, J. A., Dochherty, P., and Gersztenkorn, A., 1990, Regularization of nonlinear inverse problems: imaging the near surface weathering layer: *Inverse Prob.*, **6**, 115-131.
- Scales, J. A., Gersztenkorn, A., and Treitel, S., 1988, Fast l_p solution of large, sparse, linear systems: Application to seismic traveltime tomography: *J. Comput. Phys.*, **75**, 314-333.
- Shaw, P. R., and Orcutt, J. A., 1985, Waveform inversion of seismic refraction data and applications to young Pacific crust: *Geophys. J. Roy. Astr. Soc.*, **82**, 375-414.
- Singh, R. P., and Nyland, E., 1988, Ray bending due to strong velocity anomalies: *Geophysics*, **53**, 201-205.

- Singh, R. P., and Singh, Y. P., 1991, RAYPT-- A new inversion technique for geotomographic data: *Geophysics*, **56**, 1215-1227.
- Tanabe, 1971, Projection method for solving a singular system of linear equations and its applications: *Numer. Math.*, **17**, 203-214.
- Thurber, C. H., and Ellsworth, W. L., 1980, Rapid solution of ray tracing problems in heterogeneous media: *Bull. Seis. Soc. Am.*, **70**, 1137-1148.
- Ulrych, T. J., and Walker, C., 1982, Analytic minimum entropy deconvolution *Geophysics*, **47**, 1295-1302.
- Um, J., and Thurber, C., 1987, A fast algorithm for two-point seismic ray tracing: *Bull. Seis. Soc. Am.*, **77**, 972-986.
- Wielandt, E., 1987, On the validity of the ray approximation for interpreting delay times, in Nolet, G., Ed., *Seismic tomography with applications in global seismology and exploration geophysics*: D. Reidel Publ. Co., 85-98.
- Wiggins, R. A., 1978, Minimum entropy deconvolution: *Geoexpl.*, **16** 21-35.
- Zhou, H., 1993, Traveltime tomography with a spatial-coherency filter: *Geophysics*, **58**, 660-669.

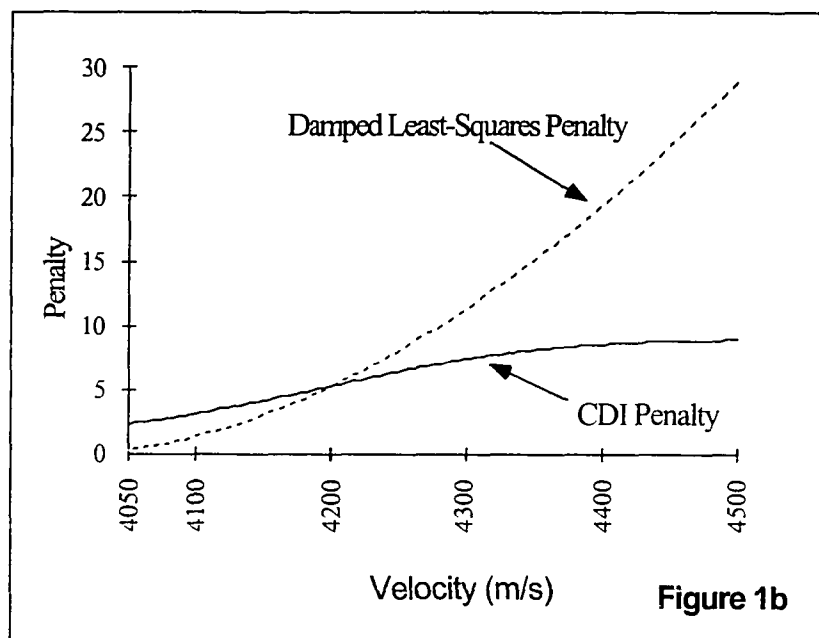
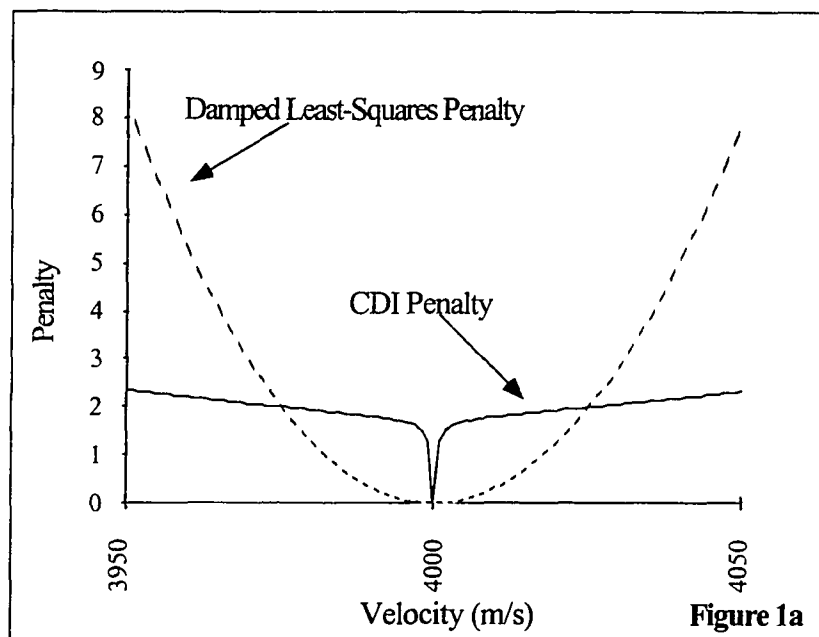


Figure 1: Comparison of CDI penalty function with damped least squares. For this example the parameters of equations (12) and (13) have the following values: $p_1 = p_2 = 1/2$, $s = 1.0E-2$, $\alpha = 3$, $c = 0.5$, $\varepsilon = 1.0E-5$. Figure 1 (a): Comparison for small parameter perturbations. Damping was adjusted for a crossover at $v = 4025$ (damping @ 912). Figure 1 (b): Comparison for moderate and large parameter perturbations. Damping was adjusted for a crossover at $v = 4200$ (damping @ 194).

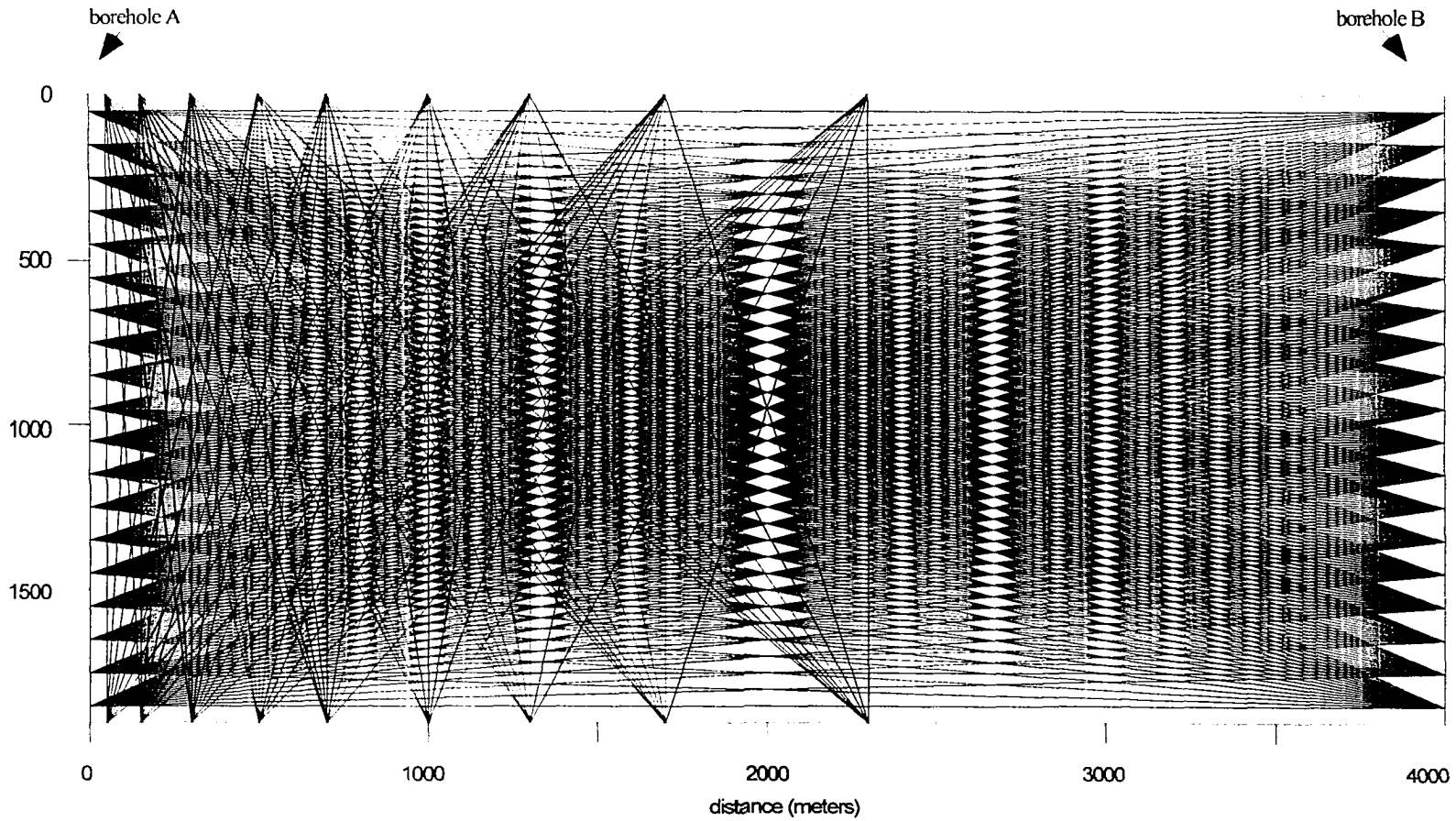


Figure 2: Raypaths used for point spread function examples. A total of 442 raypaths are used for point spread function tests. Horizontally traveling raypaths are symmetric at the anomalous point locations of Figure 3. The left side includes surface to interior raypaths which are not experimentally feasible, but demonstrate improved imaging with greater azimuthal coverage.

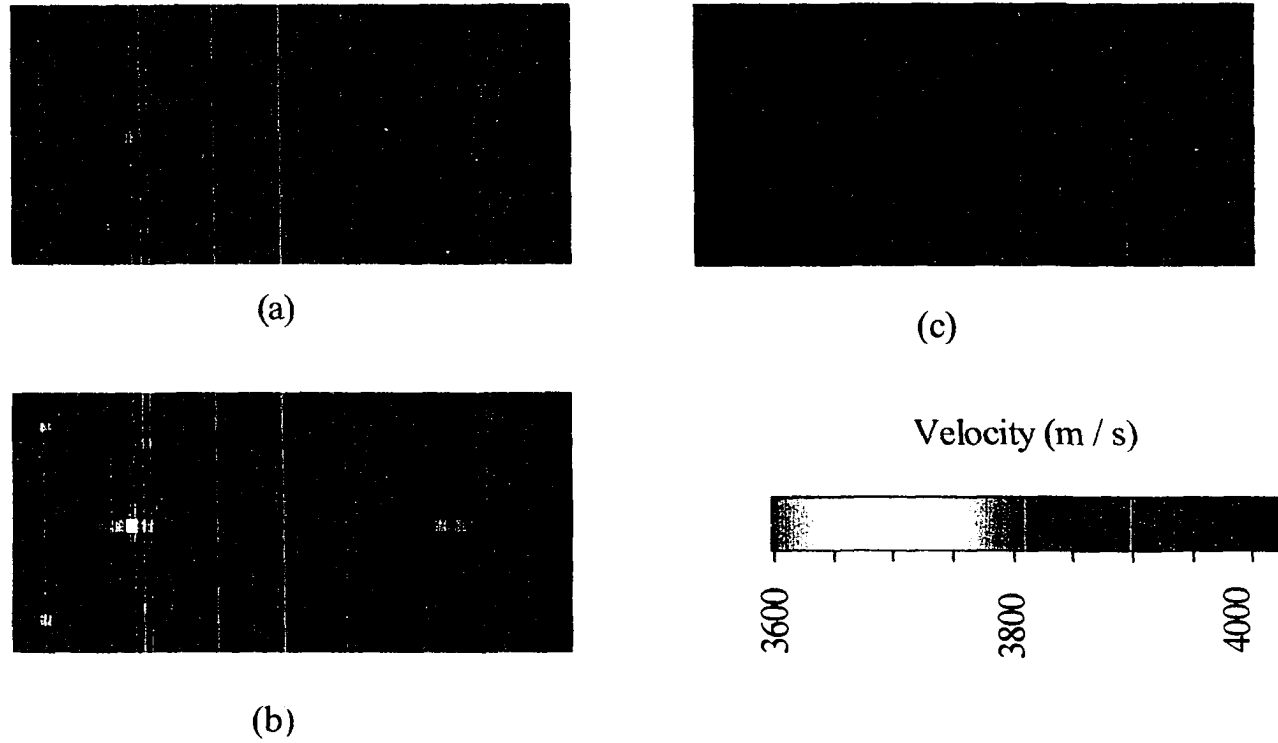


Figure 3: Point spread functions.

(a) Model velocity field; (b) Damped least-squares tomogram; (c) CDI tomogram. The damped least-square tomogram is characterized by broad, diffuse point spread functions and attenuation amplitudes. The CDI tomogram reproduces the original model to within numerical stopping criteria.

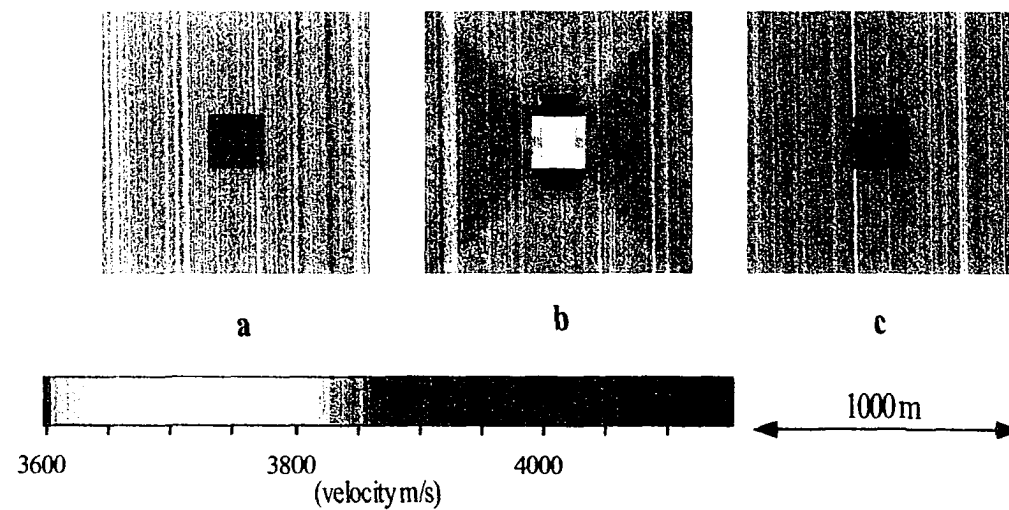


Figure 4: Brick anomaly imaging comparison.

The hypothesized velocity field is shown in (a), the damped least-squares tomogram in (b), and (c) is the CDI tomogram. The damped least-squares tomogram is attenuated and distorted while the CDI tomogram essentially matches the original velocity distribution

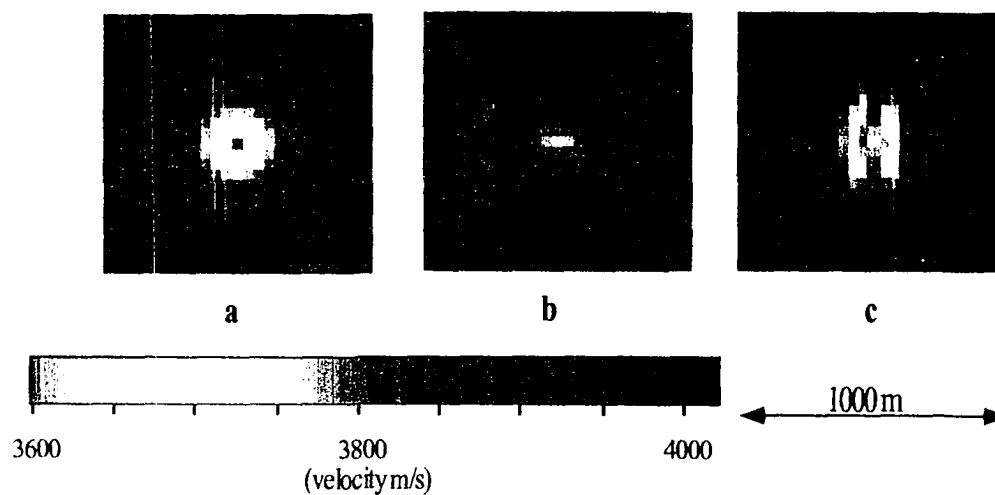


Figure 5: Blob anomaly imaging comparison.

The hypothesized velocity field is shown in (a), the damped least-squares tomogram in (b), and (c) is the CDI tomogram. Both tomograms are distorted, though smearing through the non-anomalous area is greatly reduced on the CDI tomogram. When doubling the number of raypaths, the CDI solution reproduces the original model, while the Gaussian solution only improves slightly

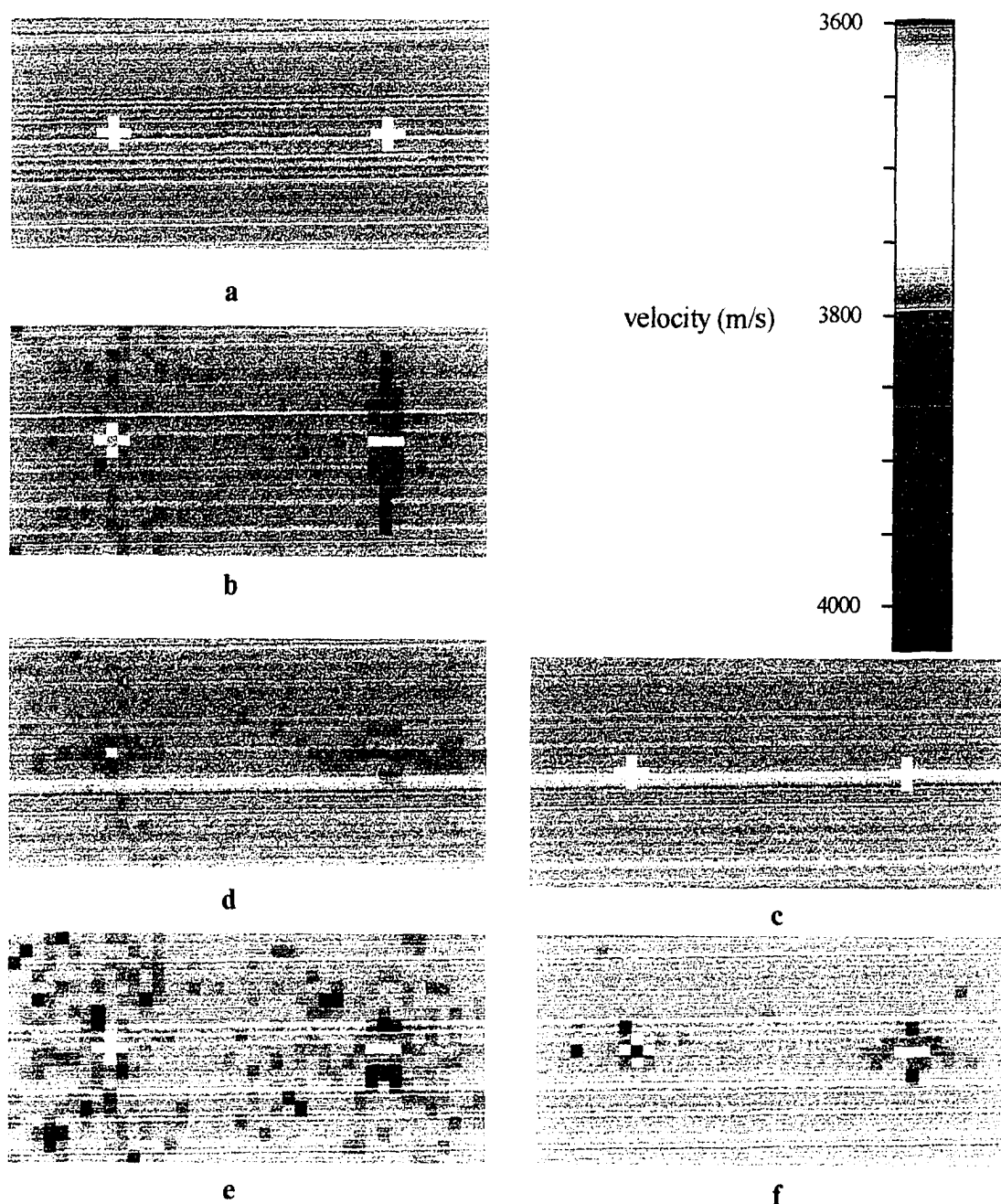


Figure 6: Plus anomaly imaging comparison. The following images are depicted: (a) Model velocity field; (b) Noise free damped least-squares tomogram; (c) Noise-free CDI tomogram; (d) Slightly overdamped noise-contaminated damped least-squares tomogram; (e) Slightly underdamped noise contaminated damped least-squares tomogram; (f) Noise-contaminated CDI tomogram. Velocity field in (a) was sampled by raypaths depicted in Figure 2. Gaussian noise was added to the synthetic data prior to inverting for (d), (e), and (f). Standard deviation of the additive noise was $1/4$ the traveltime anomaly associated with one anomalous block.

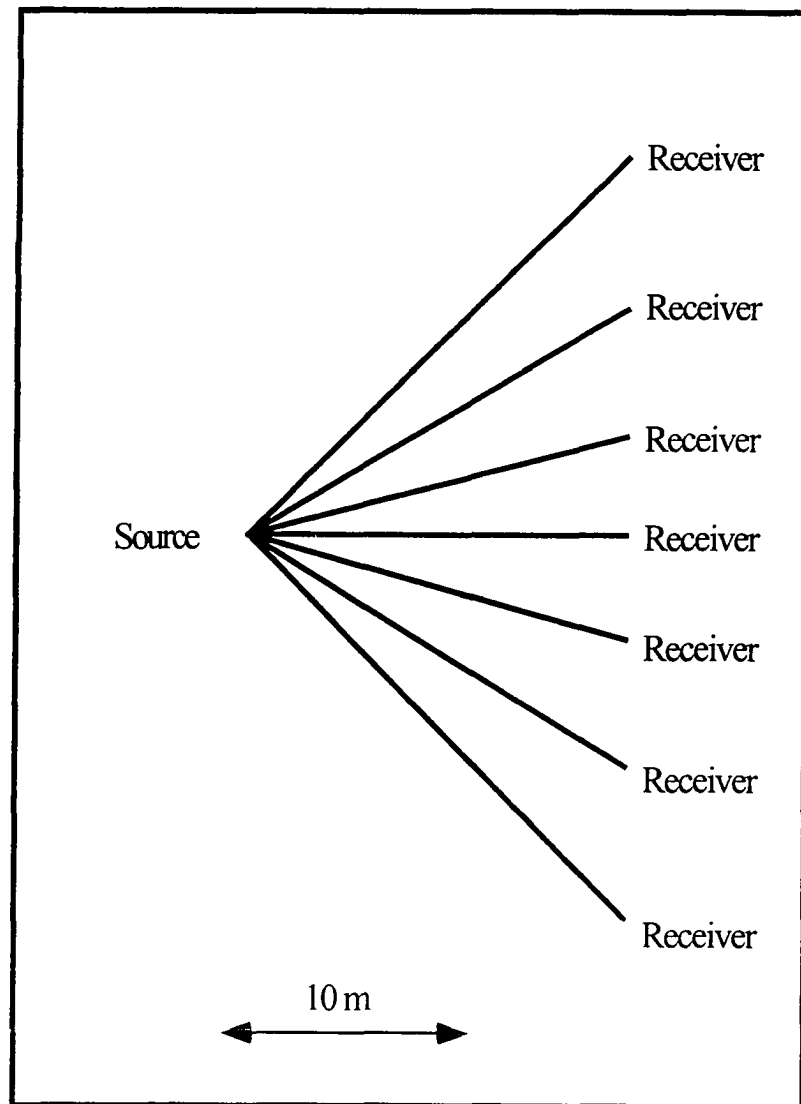


Figure 7: Azimuths used for crosshole data acquisition. Source-to-receiver azimuths used for each source location. Sources were spaced at 0.2 m intervals in the source borehole, with each source shooting into receivers located at the azimuths shown.

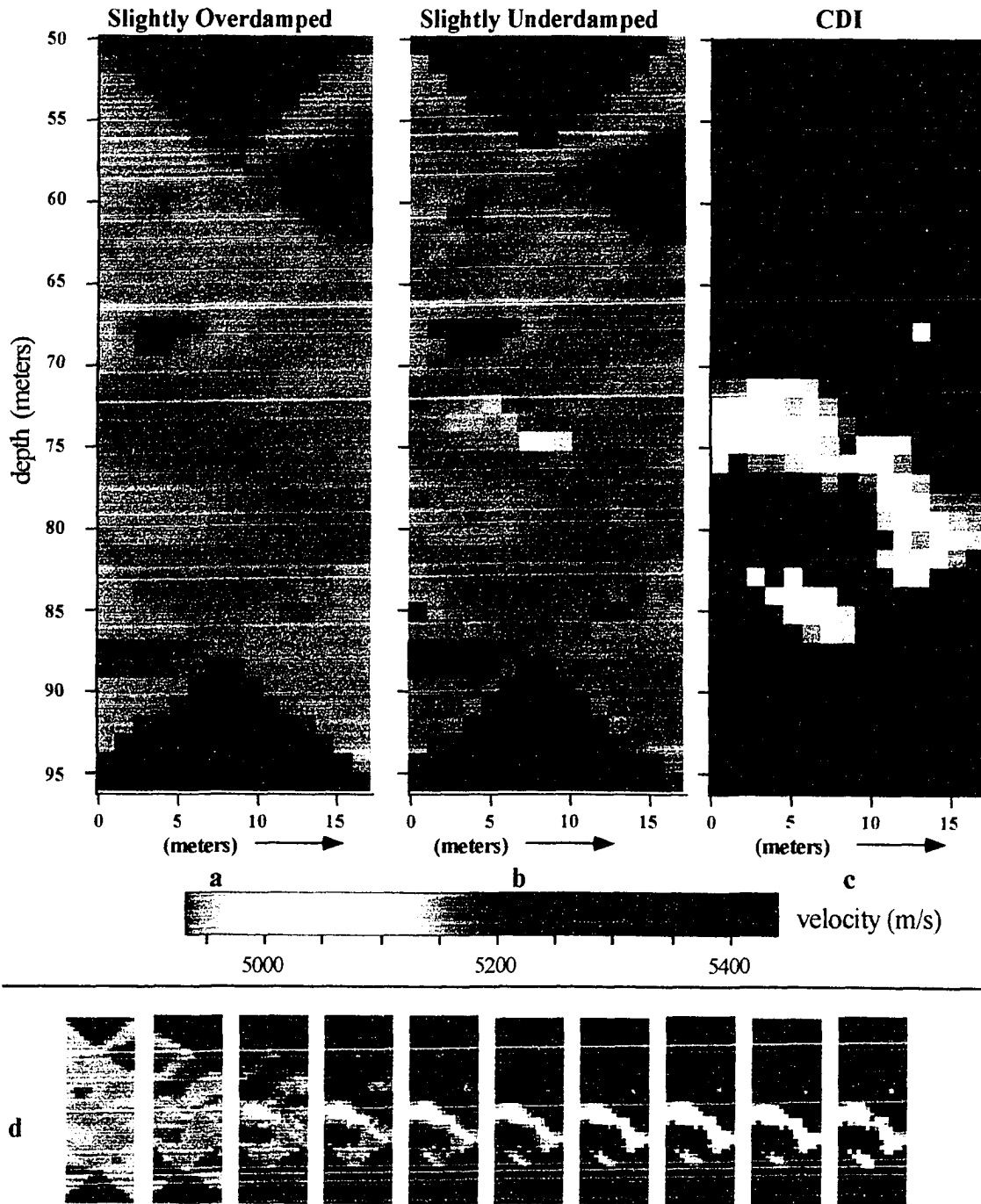


Figure 8: Tomograms based on crosshole field data: (a) Slightly overdamped least-squares tomogram; (b) Slightly underdamped least-squares tomogram; (c) CDI tomogram. The CDI tomogram provides a less attenuated, better defined anomalous zone near the center. Total anomalous area is reduced on the CDI image, and noise artifacts along acquisition azimuths are almost eliminated; (d) Intermediate tomograms produced by the CDI algorithm, shown after each of the first ten reweight iterations (eleven were required for convergence, the eleventh is shown in (c)).

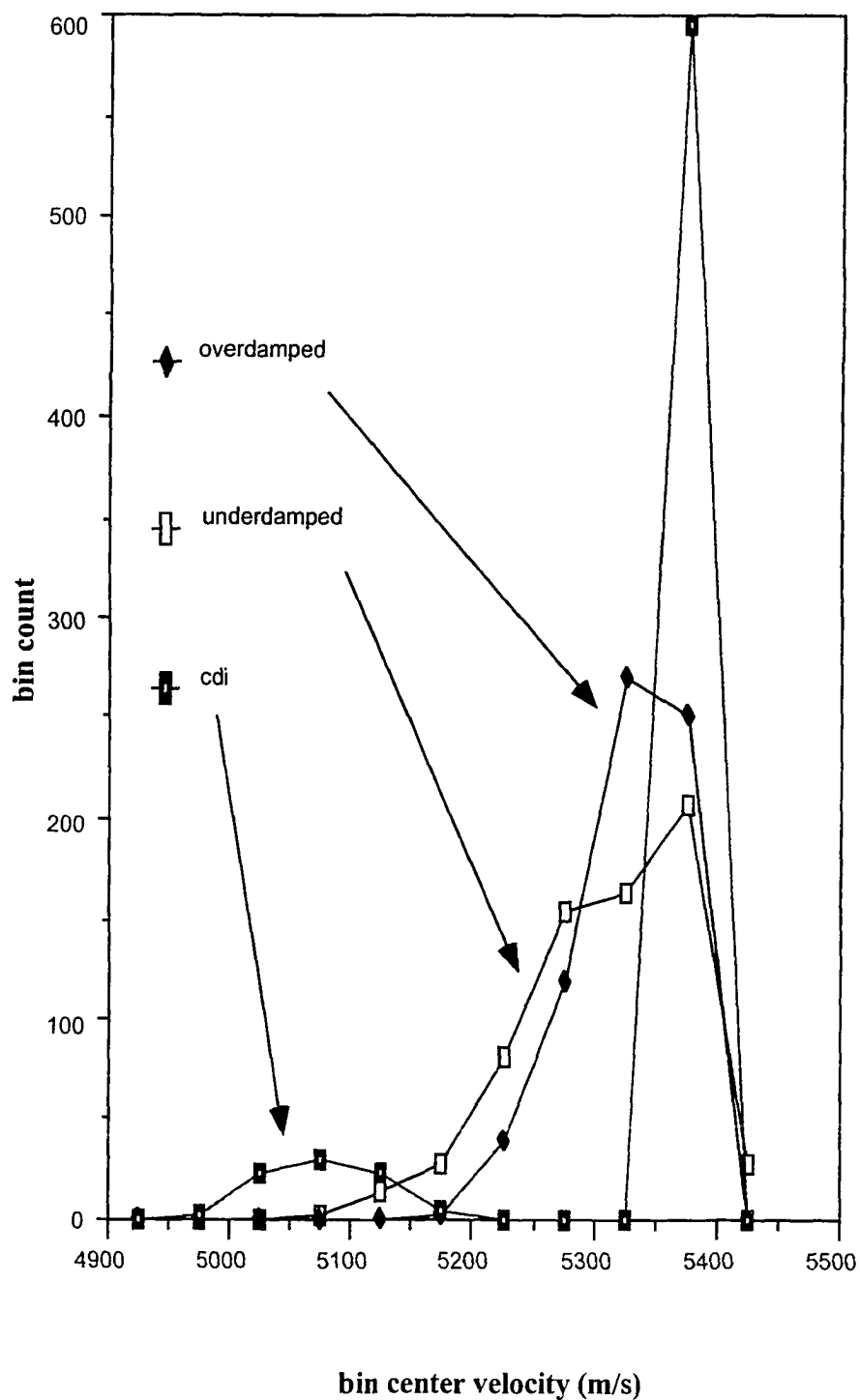


Figure 9: Comparison of estimated velocity distributions. The connected line histograms compare the empirical distributions of the estimated velocity fields shown in Figure (8a), (8b), and (8c). Notice the second low-velocity mode allowed by CDI, while damping encourages a single mode.

CHAPTER 3

Tunnel Detection Using Traveltime and Attenuation CDI

Tomography in South Korea

ABSTRACT

Crosshole seismic tomography using either first arrivals or envelope amplitudes can be used for tunnel detection. The tunnel detection problem provides an opportunity to demonstrate small, discrete, and slow-velocity anomaly imaging. Such anomalies are difficult to image tomographically, because raypath density in areas of relatively slow velocity is always poor, and the type of *a priori* information typically used to stabilize the problem discourages the existence of small, discrete, anomalies.

Crosshole seismic geotomography at most resolution scales of interest requires solution of a mixed-determined inverse problem. *A priori* information must be added to mixed-determined problems to obtain a stable, unique, solution. Most types of *a priori* information used to stabilize the inverse inhibit estimation of discrete and/or high amplitude anomalies. Composite distribution inversion (CDI) tomography was introduced to allow estimation of discrete, high-amplitude, anomalies embedded in a fairly well-known background (Clippard et al., 1995). CDI, as originally introduced, makes the *a priori* assumption of a uniformly distributed anomalous population and a normally distributed background. Additional experience with CDI now indicates that the assumption of two Gaussian populations of differing variance is intrinsically more stable

and less sensitive to user parameter choices than the original assumption of a composite Gaussian/normal population.

Reliable anomaly detection using CDI tomography requires accurate specification of a background model within which the anomalous targets are located. Source and receiver location accuracy requirements can easily exceed the accuracy of measured borehole deviation information. Systematic location errors can significantly impair estimation of the background model, particularly for small-scale engineering surveys. We show that location corrections can be derived directly from the data by matching distributions of apparent velocity or attenuation across sampling angles. We use these techniques to successfully image a tunnel near the northern border of South Korea. The combination of empirical corrections and CDI constraints are required to produce the images shown.

INTRODUCTION

Crosshole tomographic detection of small, high amplitude, anomalies, is a challenging problem. Experimental geometry restrictions limit the range of sampling angles which produces a mixed-determined inverse problem (Scales et al., 1990). A unique solution for such a problem only exists if information is added to the problem (Menke, 1984). Smoothness, flatness, or minimum solution length assumptions are most often used to stabilize the inverse, but all of these attenuate anomaly amplitudes and distribute the anomalous signal through connected systems of model parameters. Numerical investigations of crosshole tomography show that the amplitude of point anomalies are attenuated by 50% or more by damped least-squares stabilization (Menke,

1984). Imposition of *a priori* information upon the tomographic inverse also induces distinct local minima (Vasco et al., 1996). This implies that the final images are highly dependent on the starting model when downhill methods (Gill et al., 1981) are used.

Composite distribution inversion (CDI) tomography has been shown to allow detection of small, discrete, high-amplitude anomalies embedded in a low-variance background. The basic idea of CDI is to allow high amplitude anomaly estimates in truly anomalous regions while still introducing sufficient information to stabilize the inverse. This is accomplished by assuming that the model parameters being estimated are members of a composite population consisting of two sub-populations: 1) A low-variance population associated with the background parameters; and 2) A large-variance population associated with anomalous parameters. The performance of CDI tomography requires a low noise data volume and good knowledge of background parameter values. Noise associated with borehole effects and positioning errors can seriously degrade the performance of CDI tomography.

Although borehole-based seismic observations are usually contaminated by less noise than surface seismic data, several sources of noise still exist. Systematic positioning errors prevent accurate determination of the primary velocity structure, which in turn prevents detection of small anomalies using CDI. This is particularly true for small-scale engineering surveys, where small absolute errors can represent a non-negligible fraction of the borehole separation. Near-borehole effects can also strongly influence tomograms. These effects introduce imaging artifacts that originate near the borehole and ‘streak’ across the tomogram along the sampling raypaths. Parameter estimation errors occur in

other portions of the image to compensate for the ‘streaking’ errors, resulting in additional image degradation.

The objective of this paper is to demonstrate tunnel detection using CDI tomography and data-derived corrections. In the first section we briefly describe our approach to composite distribution inversion, crosshole tomography, and empirical crosshole data correction estimation. In the next section we describe the field data acquisition. We conclude with a description of processing and comparative examples of tunnel-detection using both attenuation and traveltimes tomograms.

METHODS

Seismic tomography is most commonly posed as a discrete inverse problem of the

$$\text{form} \begin{bmatrix} \mathbf{G} \\ \mathbf{D} \end{bmatrix} \Delta \mathbf{m} = \begin{bmatrix} \Delta \mathbf{d} \\ \mathbf{s} \end{bmatrix}, \quad (16)$$

where \mathbf{G} is a sensitivity matrix of partial derivatives, $\Delta \mathbf{d}$ is a vector of residuals given by $\mathbf{d}_{\text{observed}} - \mathbf{d}_{\text{predicted}}$, $\Delta \mathbf{m}$ is the vector of model parameter perturbations to be estimated, and $\mathbf{D} \Delta \mathbf{m} = \mathbf{s}$ are the constraint equations used to stabilize the inverse. Vector \mathbf{s} is usually the zero vector, but may be used to account for perturbations away from an initial model during an iterative inverse. Matrix \mathbf{G} is unavoidably large, sparse, and ill-conditioned for geotomographic problems (Lees and Shalev, 1992), which is a direct consequence of the small number of model parameters sampled by each raypath. When the constraint system is based on distributional assumptions, matrix \mathbf{D} is usually chosen to be symmetric and is related to the inverse model-parameter covariance matrix

as $\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D} = \mathbf{D} \mathbf{D}^T$. If all model parameter perturbations to be estimated are distributed as a standard normal, $n(0, \sigma^2)$, and there is no covariance between parameters, then $\mathbf{C}^{-1} = \sigma^{-2} \mathbf{I}$ and $\mathbf{D} = \sigma^{-1} \mathbf{I}$. In this case, equation (16) simplifies to the damped least-square inverse. The damped least-square inverse does not always generate a desirable solution. Often, the assumption of some covariance between model parameters is used to impose smoothing or flatness constraints (Lees and Shalev, 1992). Flatness or smoothness constraints inhibit resolution of small, discrete anomalies by attenuating the actual anomaly and distributing the total anomalous signal into smooth or flat regions of space near (hopefully) the actual anomalous area. Damped least-square solutions are usually no better. The magnitude of anomaly attenuation is quite significant. In two-dimensional problems, estimated anomaly amplitudes are often less than 50% of the true anomaly (Humphreys and Clayton, 1988; Lees and Crosson, 1989). CDI tomography was introduced to allow estimation of large amplitude, physically small, anomalies in a discrete background (Clippard and Christensen, 1995). The premise behind CDI is that the parameters to be estimated may be members of several sub-populations, each of which is described by a separate distribution function. This implies that prior assumptions used to stabilize the inverse should reflect the distribution of the composite population. For distinct anomaly detection problems, it is appropriate to assume a low-variance background component, and a high-variance anomaly component. As originally introduced, the CDI background population was assumed to be normally distributed, and the anomaly was assumed uniform. This means that no prior information

whatsoever is assumed concerning the distribution of anomalous parameters.

Computational experience indicates that less parameter sensitive results and quicker algorithmic convergence are obtained if the anomalous population is assumed to be normally distributed with large variance. Here we assume both sub-populations are normally distributed, and that there is zero covariance between model parameters.

Specifically, the background parameters are assumed to be distributed as $n(0, \sigma_b^2)$ and the anomalous parameters are assumed to be distributed as $n(0, \sigma_a^2)$. The associated probability density functions given by $f_b(\Delta m_i)$ and $f_a(\Delta m_i)$. In this case, \mathbf{D} is diagonal, with elements given by either σ_b^{-1} or σ_a^{-1} . Knowledge of whether a parameter is a member of the background or the anomaly population is usually not available in advance. This implies that it is not possible to state which elements of \mathbf{D} should be assigned the value σ_b^{-1} and which should be assigned σ_a^{-1} . We use an iterative probabilistic approach, beginning with an overly-smoothed or overly-damped estimate, to solve this problem. Given an estimate of the i^{th} parameter perturbation at the end of iteration k , $\Delta m_i^{(k)}$, the expected value for element D_{ii} of matrix \mathbf{D} during iteration $k+1$ is given by

$$D_{ii}^{(k+1)} = \frac{r f_b(\Delta m_i^{(k)}) \sigma_b^{-1} + (1-r) f_a(\Delta m_i^{(k)}) \sigma_a^{-1}}{r f_b(\Delta m_i^{(k)}) + (1-r) f_a(\Delta m_i^{(k)})}, \quad (17)$$

where r is the assumed fraction of the overall population associated with the background.

Iterative estimates of $\Delta \mathbf{m}$ are made using a sparse linear systems solver, such as LSQR (Paige and Saunders, 1982), until convergence is reached. The elements of \mathbf{D} are

updated between iterations. Convergence of this type of Picard iteration has been previously investigated (Eckhart, 1980). Computational experience indicates that this implementation of CDI tomography usually converges in three to ten iterations, and that the number of iterations required varies substantially with the choice of population distribution parameters.

CDI tomography requires a low noise level, and an accurate estimate of the reference model to image spatially small features. Although the borehole environment is intrinsically low-noise when compared with surface observations, borehole and near-borehole effects can introduce a significant level of noise. Noise associated with specific source or receiver locations generates imaging artifacts that disperse through connected systems of model parameters during the inversion. Systematic variations, such as errors associated with inaccurate borehole deviation surveys, may degrade the integrity of the derived background model. For small-scale engineering surveys, it is particularly difficult to obtain a deviation survey of sufficient accuracy. In addition, accurate knowledge of the source radiation pattern is required for attenuation tomography.

Traveltime tomography and empirical corrections

It is possible to derive both borehole deviation and radiation pattern corrections empirically from the crosshole observations. Borehole deviation corrections may be derived by first assuming that it is possible to adequately describe the actual borehole trajectory as the sum of the reported trajectory and a parametric curve associated with the deviation from the reported trajectory. Next, we note that the same approximate physical

space is sampled by all acquisition angles, and therefore the distribution of apparent velocities should be approximately the same for all angles. Any deviation in the distribution of apparent velocities is either associated with systematic borehole deviation, or variation in the earth volume sampled near the top and bottom of the survey (Figure 10). A measure of distribution equality between angular sampling partitions is then minimized as a function of the deviation trajectory parameters. If it is suspected that substantial velocity variations exist near the top or bottom of the borehole, large sample angles in these regions may be discarded from the analysis with little effect on the results. The actual process, using a low order polynomial to describe the deviation trajectory, is implemented by first writing

$$x_i(\tau) = x_{i_0}(\tau) + \sum_{j=0}^M a_j^{(i)} \tau^j, \quad (18)$$

where $x_i(\tau)$ is the actual i^{th} cartesian coordinate of some point in the borehole at a measured depth of τ , and $x_{i_0}(\tau)$ is the reported horizontal coordinate, M is the order of the polynomial, and the $a_j^{(i)}$ are coefficients to be determined. Next, the observations are partitioned into P groups of approximately equal acquisition angle. The apparent mean velocity for each observation is then computed using raypaths computed through a hypothesized initial model, $v_{\text{app}} = \frac{\eta}{t}$, where η is the raypath length and t is the observed traveltime. The overall process is fairly insensitive to the selection of the hypothetical model, since errors in the model affect raypaths from all groups in approximately the same way. Straight rays are an adequate approximation for small-scale surveys in

regions of little velocity variation. Empirical distribution functions are then assembled for each angular sampling partition, resulting in a group of distributions, , where v_{app} is apparent velocity and is a function of the $a_j^{(i)}$, with $j = 1, \dots, M$ and $i = 1, \dots, P$. A measure of overall distribution similarity is then selected as an objective function, $\Phi(F_1, \dots, F_p)$. This objective function is then maximized to determine the optimal $a_j^{(i)}$. Several measures of distribution similarity are possible (Daniel, 1990; Hogg and Craig, 1978). For some cases, a simple choice of $F(v_{app})$, such as dispersion or central tendency, is sufficient. If $F(v_{app})$ is continuous, then it is possible to differentiate the objective function with respect to each correction parameter, and directly solve for the optimal $a_j^{(i)}$ using standard methods. In this case, a system of simultaneous equations is constructed for simultaneous solution where the equation associated with the i th coordinate and the j^{th} coefficient of the system is given by

$$\frac{d}{da_j^{(i)}} \Phi(F_1, \dots, F_p) = 0. \quad (19)$$

If $F(v_{app})$ is not differentiable, which might be the case if a robust statistical measure such as the Kolmogorov-Smirnov test is used (Massey, 1951), equation (16) must be optimized using a numerical method, such as downhill simplex or simulated annealing (Nelder and Mead, 1965; Press et al., 1986).

Attenuation tomography and empirical corrections

Source radiation pattern and spreading corrections for use in attenuation tomography can be derived empirically using ideas similar to those used to derive borehole deviation corrections. Crosshole attenuation tomography is based on the equation for amplitude attenuation of a seismic wave over a region of constant propagation velocity and quality factor (Aki and Richards, 1980),

$$A_1(f, \mathbf{x}_1, \mathbf{m}) = A_0(f, \mathbf{m}) R(\mathbf{x}_0, \mathbf{x}_1) \exp \left(-\frac{\pi f |\mathbf{x}_1 - \mathbf{x}_0|}{Q(\mathbf{x}_0, \mathbf{x}_1, \mathbf{m}) c} \right) \quad (20)$$

where

- $A_1(f, \mathbf{x}_1, \mathbf{m})$ = predicted amplitude at frequency f and location \mathbf{x}_1
- $A_0(f, \mathbf{m})$ = known amplitude at frequency f and location \mathbf{x}_0
- $R(\mathbf{x}_0, \mathbf{x}_1)$ = geometrical spreading correction factor between \mathbf{x}_0 , and \mathbf{x}_1
- $Q(\mathbf{x}_0, \mathbf{x}_1, \mathbf{m})$ = effective quality factor between \mathbf{x}_0 , and \mathbf{x}_1
- \mathbf{m} = parameter vector allowing for spatial variation in Q
- c = seismic propagation velocity.

If a number of straight segments connected at N nodes is used to approximate a raypath, with node coordinates of \mathbf{x}_k , for $k = 0, N$, then equation (20) may be applied to each segment of the raypath to obtain a predicted amplitude at \mathbf{x}_N . The predicted amplitude is given by

$$A(f, \mathbf{x}_N, \mathbf{m}) = A_0 B(f, \theta) \prod_{k=1}^N R(\mathbf{x}_{k-1}, \mathbf{x}_k) \exp \left(-\frac{\pi f}{Q_k(\mathbf{m}) c_k} |\mathbf{x}_k - \mathbf{x}_{k-1}| \right) \quad (21)$$

where

- $A(f, \mathbf{x}_N, \mathbf{m})$ = predicted amplitude at \mathbf{x}_N
- A_0 = effective source amplitude at \mathbf{x}_0
- $B(f, \theta)$ = source directivity correction for frequency f and angle θ
- $R(\mathbf{x}_{k-1}, \mathbf{x}_k)$ = geometric spreading factor between points \mathbf{x}_{k-1} and \mathbf{x}_k
- $Q_k(\mathbf{m})$ = $Q(\mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{m})$, i.e. the effective Q for the k^{th} segment.

Attenuation tomography is based on inversion of observed data to match the amplitudes predicted by equation (21). The free parameters are contained in the vector \mathbf{m} . Once the parameters, \mathbf{m} , are estimated, an image of the spatial variation of Q can be constructed.

Amplitudes predicted by equation (21) are a function of frequency, implying the need to either estimate the free parameters using amplitude observations made at a single frequency, or to adopt an approach based on envelope amplitude across a larger spectrum. Q is usually not frequency dependent within the seismic bandwidth, and strongly band-limited measurements of amplitude are far more susceptible to noise than an average across the actual band of observation. This motivates an approach based on spectral ratios or shifts (McDonal et al., 1958; Quan and Harris, 1993; Quan and Harris, 1997), or an approach using envelope amplitude. The total power at a fixed frequency predicted is the square of equation (21),

$$A^2(f, \mathbf{x}_N, \mathbf{m}) = A_0^2 B^2(f, \theta) \prod_{k=1}^N R^2(\mathbf{x}_{k-1}, \mathbf{x}_k) \exp \left(-\frac{2\pi f}{Q_k(\mathbf{m}) c_k} |\mathbf{x}_k - \mathbf{x}_{k-1}| \right) \quad (22)$$

Linearization of equation (22) is accomplished by taking the natural logarithm of both sides of equation, which yields

$$\left. \begin{aligned} \text{Log}(A^2(f, \mathbf{x}_N, \mathbf{m})) &= 2 A_c - 2\pi f \sum_{k=1}^N \frac{|\mathbf{x}_k - \mathbf{x}_{k-1}|}{Q_k(\mathbf{m}) c_k} \\ \text{where} \\ A_c &= \text{Log}(A_0) + \text{Log}(B(f, \theta)) + \text{Log}(R(\mathbf{x}_0, \mathbf{x}_1)) \\ R(\mathbf{x}_0, \mathbf{x}_N) &= \prod_{k=1}^N R_s(\mathbf{x}_{k-1}, \mathbf{x}_k) \end{aligned} \right] \quad (23)$$

The term A_c refers to a group of factors that may be thought of as a lumped correction for the raw data. $R(\mathbf{x}_0, \mathbf{x}_N)$ is the effective spreading factor for the entire raypath. Since $Q \neq Q(f)$ (Quan and Harris, 1997), equation (23) may be integrated over some finite band of frequencies, $\Delta f = f_1 - f_0$, in an arrival envelope to produce

$$\left. \begin{aligned} E(\mathbf{m}) &= \int_{f=f_0}^{f_1} \text{Log}(A^2(f, \mathbf{x}_N, \mathbf{m})) df = 2 A'_c(\theta) - \sum_{k=1}^{n_p} \alpha_k(\mathbf{m}) |\mathbf{x}_k - \mathbf{x}_{k-1}| \\ \text{where} \\ A'_c(\theta) &= \Delta f \text{Log}(A_0) + \Delta f \text{Log}(R(\mathbf{x}_0, \mathbf{x}_1)) + \int_{f=f_0}^{f_1} \text{Log}(B(f, \theta)) df \\ \text{and} \\ \alpha_k(\mathbf{m}) &= \frac{\pi(f_1^2 - f_0^2)}{Q_k(\mathbf{m}) c_k} \end{aligned} \right] \quad (24)$$

Equation (24) is the log of observed power integrated over a finite frequency band, given a set of model parameters and a raypath. An objective function can be directly defined that uses this quantity to compare predictions with observations. The tomography

problem is then to estimate the model parameters that produce the best match between observations and predictions.

The term $\alpha_k(\mathbf{m})$ in equation (24) is the effective attenuation exponent for the k^{th} raypath segment. The spatially varying attenuation exponent, $\alpha_k(\mathbf{x}, \mathbf{m})$, can be used as the fundamental quantity imaged in attenuation tomography. In this case, the model parameters are selected to describe the spatial variation of $\alpha_k(\mathbf{x}, \mathbf{m})$, rather than Q itself. Given an estimate of the attenuation exponent, the frequencies of the observation, and the propagation velocity, a spatially varying Q can be directly computed.

To adequately describe the spatial variation of the attenuation exponent, many free model parameters are usually required. Only a small fraction of the parameters affect any single sample for most commonly used parameterizations. This results in a large, sparse, inverse problem, which requires addition of *a priori* information to have a unique, stable solution. Both least-squares and CDI *a priori* information are stated in terms of the probabilities of perturbation magnitudes. This motivates writing equation (24) in terms of a set of background model parameters and perturbations, $\alpha = \alpha(\mathbf{x}, \mathbf{m}_0 + \Delta\mathbf{m})$, where \mathbf{m}_0 is an initial vector of parameters and $\Delta\mathbf{m}$ represents the perturbations that are actually estimated during inversion. Using perturbations, equation (24) may be used to specify \mathbf{G} and $\Delta\mathbf{d}$ in equation (16):

$$\left. \begin{aligned} G_{ij} &= - \sum_{k=1}^{N^{(i)}} \frac{\partial}{\partial m_j} \alpha_k^{(i)}(\mathbf{m}) \\ \Delta d_i &= E_0^{(i)} - E^{(i)}(\mathbf{m}_0) \\ E_0^{(i)} &= \text{observed envelope power for ray } i \text{ over frequency band } (f_0, f_1) \end{aligned} \right\} . \quad (25)$$

Equations (16) and (25), and the terms contributing to $A_c'(\theta)$, allow iterative estimates of $\mathbf{m}^{(q)} = \mathbf{m}^{(q-1)} + \Delta \mathbf{m}^{(q-1)}$ using a standard sparse linear systems solver such as LSQR (Paige and Saunders, 1982) or SIRT (Gilbert, 1972; Ivansson, 1986). This only leaves the task of determining the terms of $A_c'(\theta)$ before attenuation tomograms can be computed using observed envelope amplitudes.

The terms of $A_c'(\theta)$ are only approximately known prior to the experiment. Image improvement is possible if empirical corrections are used to refine initial estimates of the parameters to achieve maximum consistency with the observations prior to inversion. The geometrical spreading factor can be approximated using a simple correction for spherical divergence. The effective source amplitude is determined by considering only raypaths at the maximum amplitude sampling-angle, usually $\theta \cong 0$. Source directivity corrections are then derived in a manner similar to that used for borehole deviation corrections in the previous section. The term $B(f, \theta)$ may be considered as a product of known source directivity function and a correctional factor,

$$B(f, \theta) = \lambda(\theta) B_0(f, \theta) , \quad (26)$$

where $\lambda(\theta)$ is smoothly varying. This allows writing

$$A_c'(\theta) = \Delta f \text{Log}(A_0) + \Delta f \text{Log}(R(\mathbf{x}_0, \mathbf{x}_1)) + \text{Log}(\lambda(\theta)) + \int_{f=f_0}^{f_1} \text{Log}(B(f, \theta)) df$$

The term $\int_{f=f_0}^{f_1} \text{Log}(B(f, \theta)) df$ can be computed using known, approximate, source

directivity information. If no such information is available, $B(f, \theta)$ is assumed to be unity. In this case, the integral annihilates and all frequency and/or angular source variation is compensated for in the empirically derived correctional term, $\text{Log}(\lambda(\theta))$.

$\text{Log}(\lambda(\theta))$, is derived by first partitioning the data into groups of common acquisition angle and noting that all groups sample the same physical space. This means that the empirical distribution of apparent attenuation exponents for all angular sampling partitions should be the approximately equal. The same procedure used to minimize the difference of apparent velocity distributions (above) is then used to estimate the empirical correction factors for each group of acquisition angles.

Another issue in attenuation tomography is the question of whether intrinsic or scattering Q dominates the observed amplitude decay. Equation (20) states that Q related amplitude decay is a simple exponential function of the distance over which energy propagates. This holds for intrinsic Q , but not for scattering. In layered systems, scattering Q can be a strongly nonlinear function of raypath incidence angle and frequency. In this case, scattering Q can be important in controlling the actual amplitudes observed (Neep et al., 1996). If scattering Q is assumed to be random and homogeneous throughout an imaging volume, which is possible for non-layered systems, scattering Q

and intrinsic Q can be lumped into a single quantity. The raypaths for our investigation presumably propagated through uniform granite, and thus specific treatment of scattering Q is unnecessary.

DATA ACQUISITION

Our experimental objective was to investigate the feasibility of tunnel detection using crosshole tomography (Figure 11). A two-meter diameter tunnel, in South Korea, is known to be located approximately midway between two boreholes at a subsurface depth of approximately seventy-five meters. The boreholes are located fifteen meters apart at the surface, and both boreholes have a measured deviation of approximately one meter at the bottom hole location, yielding a bottom hole separation of approximately sixteen meters. The lithology at the acquisition site is granite, and there is no evidence indicating lithologic variation or heterogeneity at the scale of seismic wavelengths in the survey area. Here, we present one example of a series of experiments.

A custom electrode-ignited gas-filled bladder source was constructed for tunnel detection experiments (Figure 12). The source has a bandwidth of 300-7000 Hz, with a dominant frequency of approximately 1500 Hz. Data repeatability is excellent (Rechtien et al., 1993).

The receiver was a high frequency hydrophone with a bandwidth that significantly exceeded that of the source. A time-domain sampling interval of $1 \mu\text{sec}$ was used to allow very precise picking and identification of high frequency noise components. The

associated Nyquist frequency of 50 kHz greatly exceeds that required to adequately sample the source bandwidth.

Data are acquired in a series of seven acquisition runs, each of which is conducted at a fixed sampling angle. A single source and receiver are used during each run, and are maintained at a fixed vertical separation for all source and receiver locations. A total of seven runs, representing sampling angles between -45 degrees and $+45$ degrees, are used (Figure 13). At the beginning of each run, both the source and receiver are lowered to the deepest sampling depth. Data are then acquired at 0.2 meter vertical intervals as the source and receiver are simultaneously raised. A total of 150 samples are collected during each run, producing 1050 unique raypaths. The fully sampled depth range extends from approximately sixty meters to ninety meters, subsurface. Each square meter of the fully sampled area is crossed by approximately forty separate raypaths.

DATA ANALYSIS AND RESULTS

The first task of tomographic imaging is to determine an appropriate background reference model. To do this, we use reported well deviations and the straight raypath approximation to compute an empirical distribution of apparent mean velocities. The maximum likelihood (modal) point of the distribution is selected as a constant velocity approximation to the target area. The median or mean lead to smaller prediction error than the mode, but a smaller fraction of the model space can be accurately imaged using zero perturbation away from the reference model. Since distribution assumptions employed to stabilize the inverse are stated in terms of perturbation density functions, the

mode is a superior choice for selecting a constant-valued background model. An apparent mean attenuation exponent is determined in a similar way.

Selection of a model-space parameterization is the next step. An appropriate parameterization must allow adequate description of the imaging objective, but should not use more parameters than necessary to do so. Square, constant velocity basis functions are appropriate for the purpose of detecting small, discrete anomalies. Square basis functions are associated with spatially restricted parameters, which is desirable when imaging anomalies with sharp boundaries. Another advantage of square basis functions is the simplicity of traveltimes and partial derivative computation, given a raypath. The primary disadvantage is that the resulting velocity model is discontinuous, and raytracing methods requiring velocity-gradient information must use approximations that are fundamentally inconsistent with the model itself. Application of Snell's law, rather than the ray equation, is consistent with a square basis function model, but artifacts caused by the geometry of the boundaries produce far more spurious results than those associated with gradient approximations.

Square basis functions used with a dimension of one meter because the basis function dimension should be no larger than half the diameter of the imaging target. This ensures that at least one parameter will represent a space entirely within the anomalous region. Smaller basis functions would needlessly increase the condition number of the system in an attempt to gain resolution at a scale smaller than is required to detect the anomaly. The imaging area is selected to include the uppermost and lowermost source/receiver

locations, though the central area of the uppermost and lowermost fifteen meters are not sampled by all acquisition angles.

Using the parameterization described above, a series of damped-least-squares traveltimes and attenuation tomograms were computed using a suite of damping values ranging from considerably overdamped to somewhat underdamped. Both straight and bent ray algorithms were tested, with only very minor differences, and only straight ray results are shown in this paper. The critically damped tomogram is selected using subjective judgement (Figure 14). A low velocity zone near the center of the tomogram is observed, but many other anomalous features are also present. Some of the anomalous signal is apparently acquisition artifact related, and precise interpretation of a tunnel location is not possible. An initial damped least-square attenuation tomogram is shown in Figure 15. As expected, the attenuation tomogram is more strongly influenced by near-borehole effects than the traveltimes tomogram. Several 'streaks' begin near the borehole and traverse the image along acquisition azimuths, indicating the existence of pronounced near-borehole effects.

The next task is to derive empirical borehole deviation corrections using equation (18) and minimization of empirical velocity distribution differences. Initially, a first order polynomial describing a linear deviation in the vertical plane between boreholes is used. Since the surface coordinates are known, the constant term is not considered a free parameter. Several test statistics were tried as optimization criteria, all of which produced nearly identical results. Related, but distinct graphical diagnostics of effectiveness indicate improvement in consistency of observed velocity between

partitions. One such diagnostic is to compute a measure of median velocity dispersion between angular partitions as a function of perturbation away from the derived optimal parameters (Figure 16). A well-defined minimum is present at the optimal deviation indicated by distribution matching. This is a qualitative indication of an acceptable solution. A comparison of the median velocities of each group before and after corrections is shown in Figure 17.

Corrections derived using second order terms resulted in very small second order coefficients, which suggests that wellbore curvature is negligible over the short observation interval. Derivation of two-dimensional corrections also yielded no improvement over the one-dimensional first order corrections. The correction adopted is linear and one-dimensional. The corrections bottom hole separation is 16.8 m, as opposed to the reported separation of 16 m. This corresponds to a five percent decrease in slowness at the bottom of the imaging area. The geometrical spreading term $R(\mathbf{x}_0, \mathbf{x}_N)$ is calculated using the empirically derived borehole derivation corrections. Terms A_0 , and $\lambda(\theta)$ are re-derived for the seven angular partitions. Tomograms are computed using the empirical deviation corrections, and improvement is noticeable.

The next step is to apply CDI to quell dispersion of anomalous signal through numerically connected systems of parameters, and allow large perturbation values in truly anomalous regions. Highly overdamped versions of both the attenuation and traveltimes images are selected as the initial models. CDI can be parameterized to yield a standard least-squares result or a tightly constrained result, depending on the choice of *a priori* population fractions and assumed parameter variance. At this time, no objective criteria

are available for determination of the appropriate fractions and variances, and the intrinsic non-uniqueness of geotomography allows a wide variety of images that adequately satisfy observations. This makes the imaging process interpretive, as is the case when using smoothing, flatness, or damped least-squares. The general characteristics are similar to those of standard constraint methods in that as more information is added images become more regularized, but the portions of the image that receive the constraints can be quite different. Tightly constrained solutions have more prediction error and less noise; loosely constrained images have less prediction error and more noise. Large values of background population fraction and low values of background population tend to attenuate and distribute anomalies. Large variance values selected for the anomalous fraction allow small anomalous areas to receive large anomaly perturbation estimates. The specific choice of parameters is guided by the imaging objective, the prediction error associated with the assumption, and interpretive discretion.

Considerable variation in what could be considered a final CDI tomograms is possible. Two choices of attenuation CDI tomograms are shown in Figure 18. Near borehole effects are still pronounced, though ‘tightening’ the CDI constraints highlights dominant anomalous features while suppressing many artifacts. Location specific corrections, or a less noise-sensitive inversion procedure (Bube and Langan, 1997), might improve the image, but neither are used here. Traveltime tomography is less sensitive to near-borehole effects than attenuation tomography, and ‘tight’ CDI inversion produces a tomogram with no anomalies other than the tunnel itself (Figure 19). This did come at the expense of increased prediction error, and it should be kept in mind that a continuum

between damped least-squares inversion and 'tight' CDI is possible. It is not possible, however, to suppress noise while simultaneously allowing the large amplitude estimates possible with CDI using damped least-squares inversion. A comparison of an underdamped tomogram with the 'tight' CDI tomogram using the same color scales is shown in Figure 18. The empirical corrections provide considerable improvement when compared with the image in Figure 14. Still, the anomaly is highly attenuated, and noise is starting to dominate signal.

CONCLUSIONS

Crosshole tomographic detection of a known tunnel in South Korea is demonstrated using both attenuation and traveltimes tomography. CDI tomography allows imaging of small, discrete, anomalies in a well-known background. Accurate estimation of the background model parameters can be impaired by systematic noise, such as inaccurate borehole deviation surveys or source directivity functions. A method of deriving empirical borehole deviation and source directivity corrections using sampling assumptions and distribution matching is introduced. These corrections improve image quality, and are required to achieve the results shown. The choice of CDI parameters for a specific problem is interpretive.

ACKNOWLEDGEMENTS

Funding for this project was provided in part by National Science Foundation grants EAR91-18090, EAR-94-05471, and EAR-95-06379 and United States Geological Survey grant 14-08-0001-A0574, which was administered through the Alaska Volcano Observatory. Scholarship support was provided by the Alaska Geophysical Society and Exxon through the SEG Scholarship Foundation.

We thank Robert F. Ballard for his assistance in making the field data available for publication. Field data were collected under the Tunnel Detection Program sponsored by the U.S. Army Belvoir Research, Development, and Engineering Center, Fort Belvoir, Virginia, and the U.S. Waterways Experiment Station, Vicksburg, Mississippi. Permission was granted by the Chief of Engineers to publish this data. The views of the authors do not purport to reflect the position of the Department of the Army or the Department of Defense.

REFERENCES

- Aki, K., and Richards, P., 1980 Quantitative Seismology: W. H. Freeman and Company.
- Bube, B. P., and Langan, R. T., 1997, Hybrid l_1, l_2 minimization with applications to tomography: *Geophysics*, **62**, 1183-1195.
- Clippard, J. D., and Christensen, D. H., 1995, Composite distribution inversion applied to crosshole tomography: *Geophysics*, **60**, 1283-1294.

- Daniel, W. D., 1990, Applied nonparametric statistics: PWS-Kent Publishing.
- Eckhart, U., 1980, Weber's problem and Weisdeld's algorithm in general spaces: Math. Programming, **18**, 186-196.
- Gilbert, P., 1972, Iterative methods for the three-dimensional reconstruction of an object from projections: J. Theor. Biol., **36**, 105-117.
- Gill, P. E., Murray, W., and Wright, M. H., 1981, Practical optimization: Academic Press.
- Hogg, R. V., and Craig, A. T., 1978, Introduction to Mathematical Statistics, 4th edition: Macmillan Publishing Co., Inc.
- Humphreys, E., and Clayton, R. W., 1988, Adaptation of back projection tomography to seismic travel time problems: J. Geophys. Res., **93**, 1073-1085.
- Ivansson, S., 1986, Seismic borehole tomography - Theory and computational methods: Proc. IEEE, **74**, 328-338.
- Lees, J. M., and Crosson, R. S., 1989, Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data: J. Geophys. Res., **94**, 5716-5728.
- Lees, J. M., and Shalev, E., 1992, On the stability of P-wave tomography at Loma Prieta: A comparison of parameterizations, linear and nonlinear inversions: Bull. Seis. Soc. Am., **82**, 1821-1839.
- Massey, F. J., 1951, The Kolmogorov-Smirnov test for goodness of fit: J. Amer. Statist. Assoc., **46**, 68-78.
- McDonal, F. J., Angona, F. A., Mills, R. L., Sengbush, R. L., Van Nostrand, R. G., and White, J. E., 1958, Attenuation of shear and compressional waves in Pierre shale: Geophysics, **23**, 421-439.

- Menke, W., 1984, *Geophysical data analysis: Discrete inverse theory*: Academic Press Inc.
- Neep, J. P., Sams, M. S., Worthington, M. H., and O'Hara-Dhand, K. A., 1996, Measurement of seismic attenuation from high-resolution crosshole data: *Geophysics*, **61**, 1175-1188.
- Nelder J. A., and Mead, R., 1965, A simplex method for function minimization: *Computer J.*, **7**, 308-313.
- Paige, C. C., and Saunders, M. A., 1982, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.* **8**, 43-71.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., 1986, *Numerical Recipes*: Cambridge, Univ. Press.
- Rechtien, R. D., Hambacker, K. L., and Ballard Jr., R. F., 1993, A high frequency sparker source for the borehole environment: *Geophysics*, **58**, 660-669.
- Quan, Y., and Harris, J. M., 1993, Seismic attenuation tomography based on centroid frequency shift: 63rd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 41-44.
- Quan, Y., and Harris, J. M., 1997, Seismic attenuation tomography using the frequency shift method: *Geophysics* **62**, 895-905.
- Scales, J. A., Dochherty, P., and Gersztenkorn, A., 1990, Regularization of nonlinear inverse problems: imaging the near surface weathering layer: *Inverse Prob.*, **6**, 115-131.
- Vasco, D. W., Peterson Jr., J. E., and Major, E. L., 1996, Nonuniqueness in traveltime tomography: Ensemble inference and cluster analysis: *Geophysics*, **61**, 1209-1227.

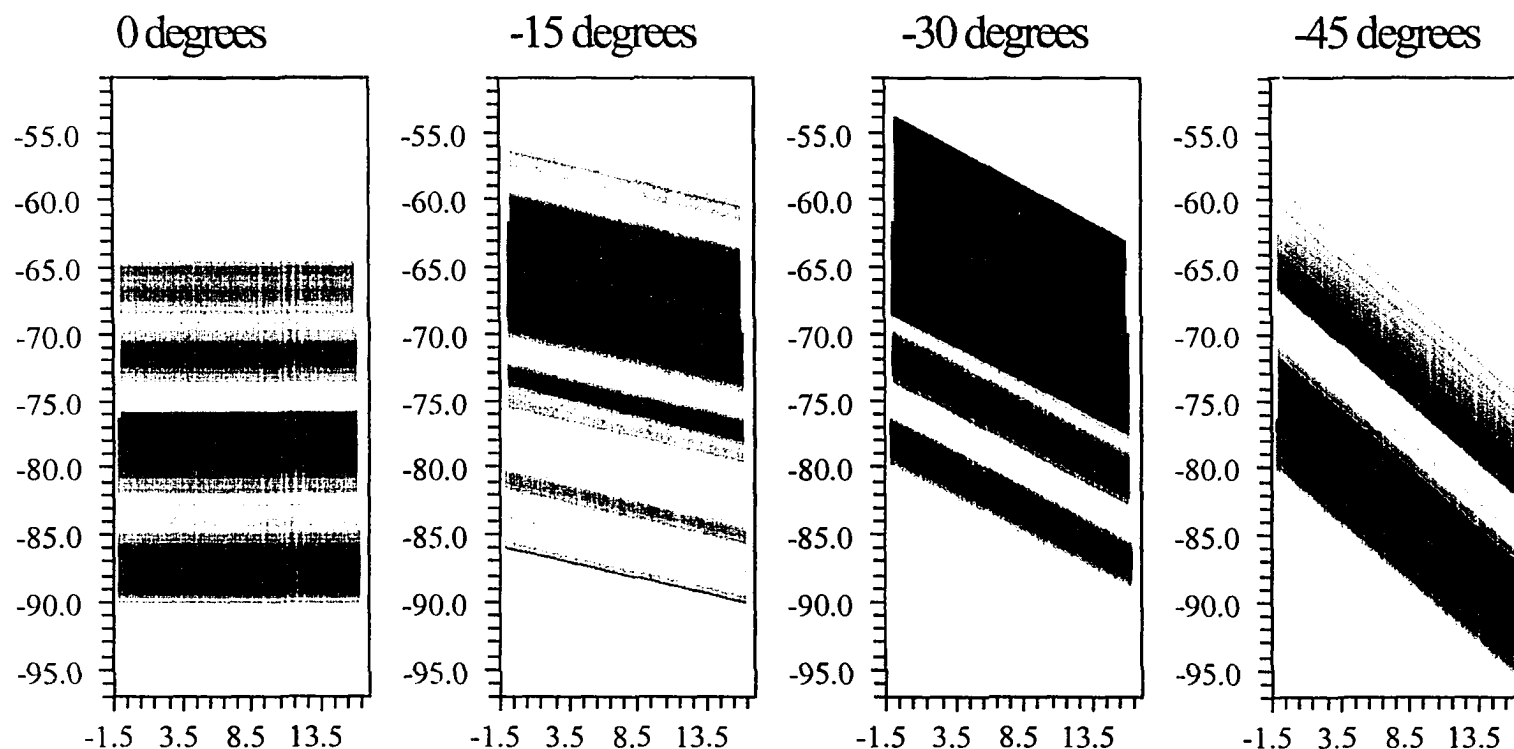


Figure 10: Differences in area sampled by different acquisition angles. Data are acquired at a small number of acquisition angles. Each suite of acquisition angles samples approximately the same area, with differences occurring near the top and bottom of the survey. Differences in the empirical distributions of apparent velocities for different acquisition angles are associated with systematic errors in raypath length calculation, or the small differences in the area sampled. Inaccurate borehole deviation surveys result in systematic raypath length calculation errors that are a function of sampling angle. Straight rays and varying grayscale are used to indicate relative apparent mean velocities observed for acquisition angles zero to negative forty-five degrees.

Scale Sketch

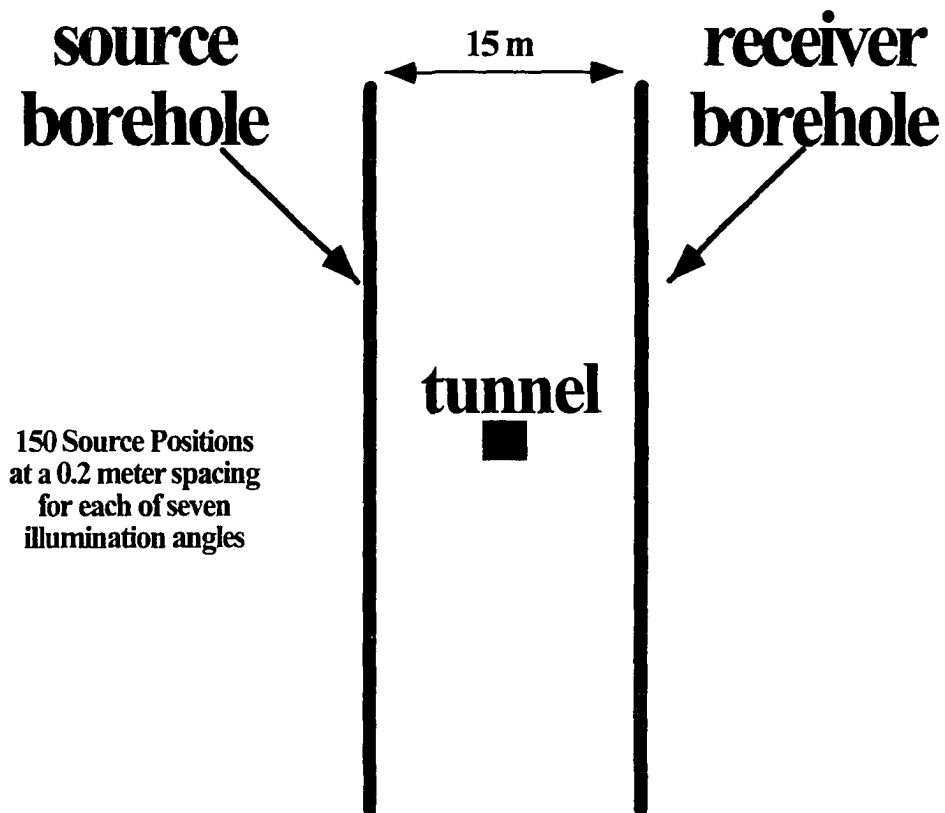


Figure 11: Diagram of experimental objective.

SOURCE

High frequency spark 300 - 7000 Hz, $f_0 = 1500$ Hz

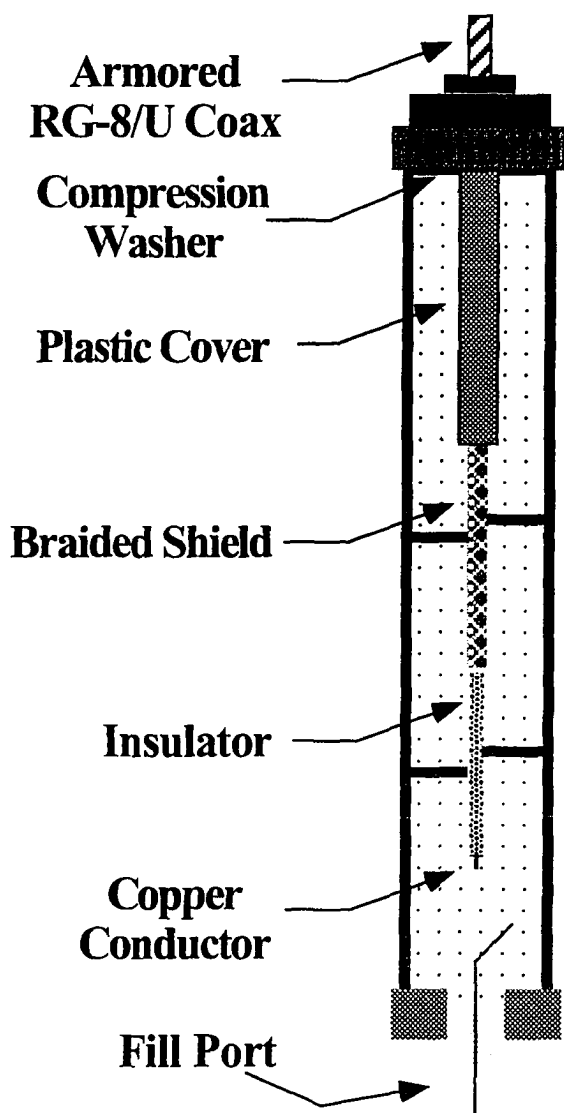


Figure 12: Arc-discharge source used during the experiment.

Illumination Angles

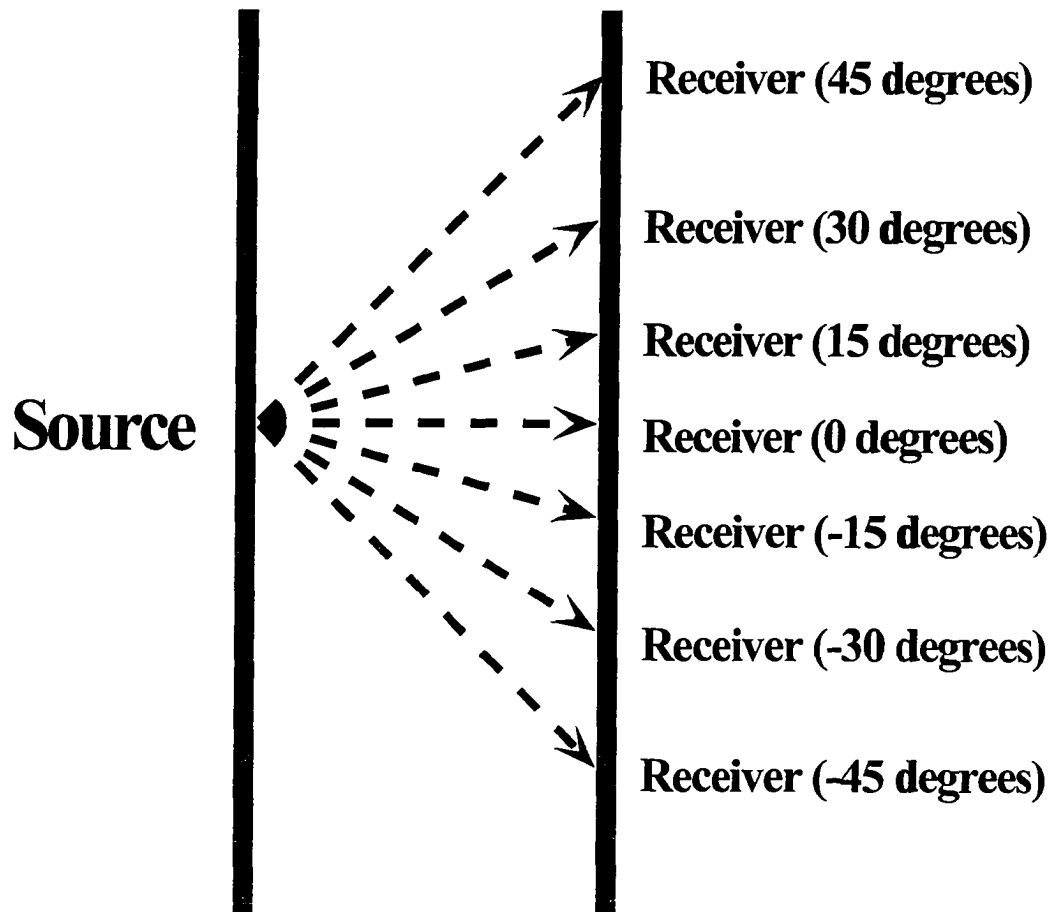
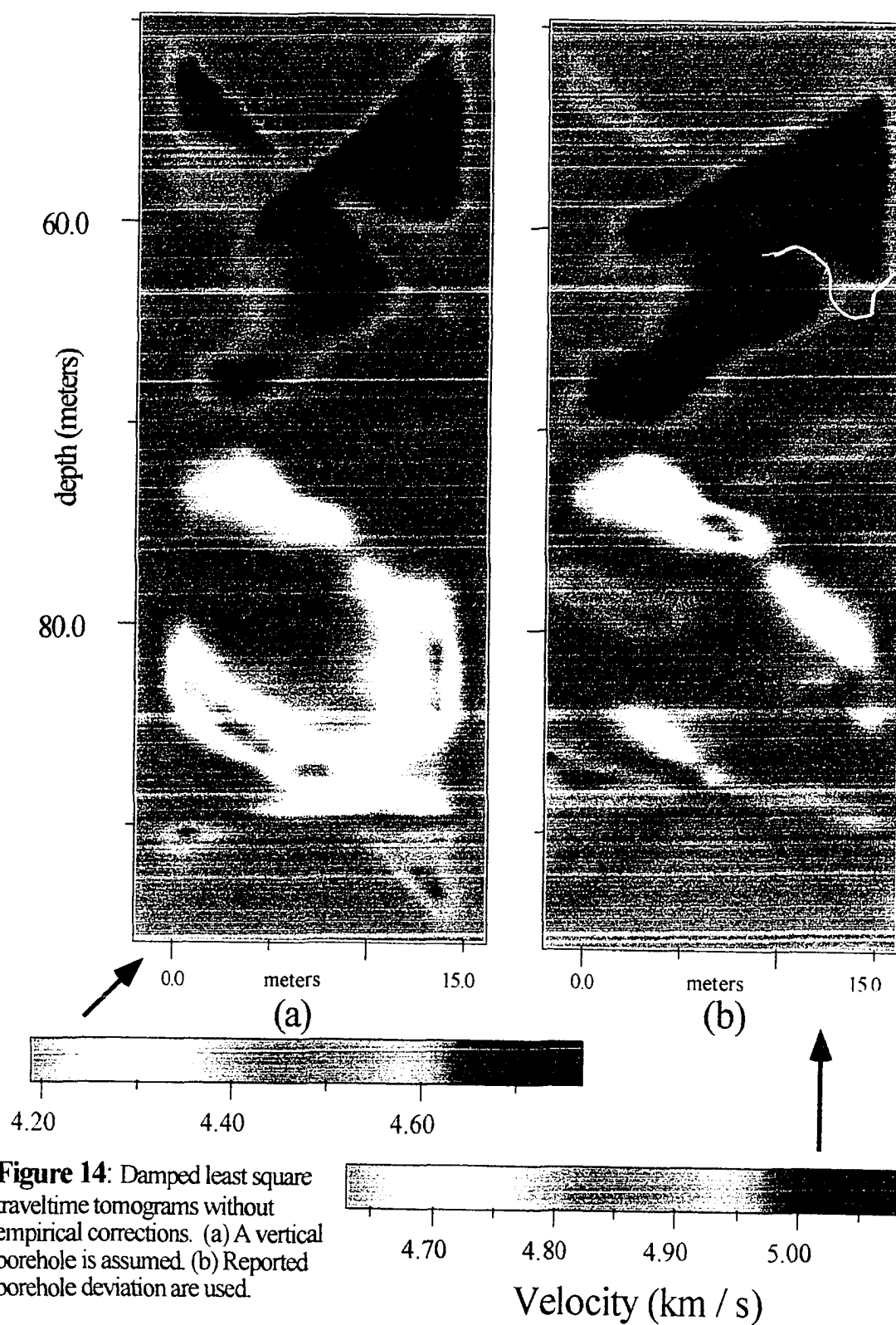


Figure 13: Crosshole sampling angles. Seven sampling angles are used with 150 source-receiver positions for each angle. Each source and each receiver location in the primary survey area is sampled by all seven angles. The vertical spacing between both shot and receiver points is 0.2 meters.



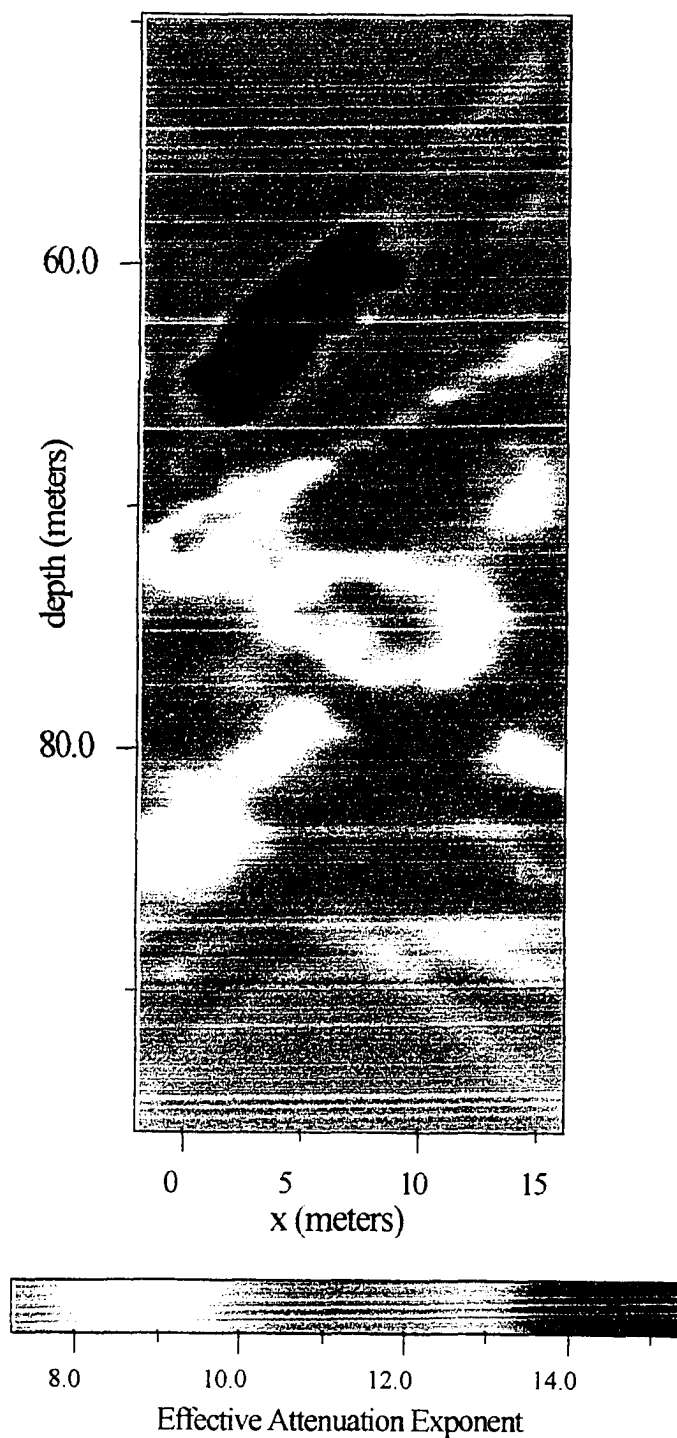


Figure 15: Damped least square attenuation tomogram without empirical corrections. The tomogram is based on reported borehole deviations only. Streaking associated with near-borehole effects is apparent.

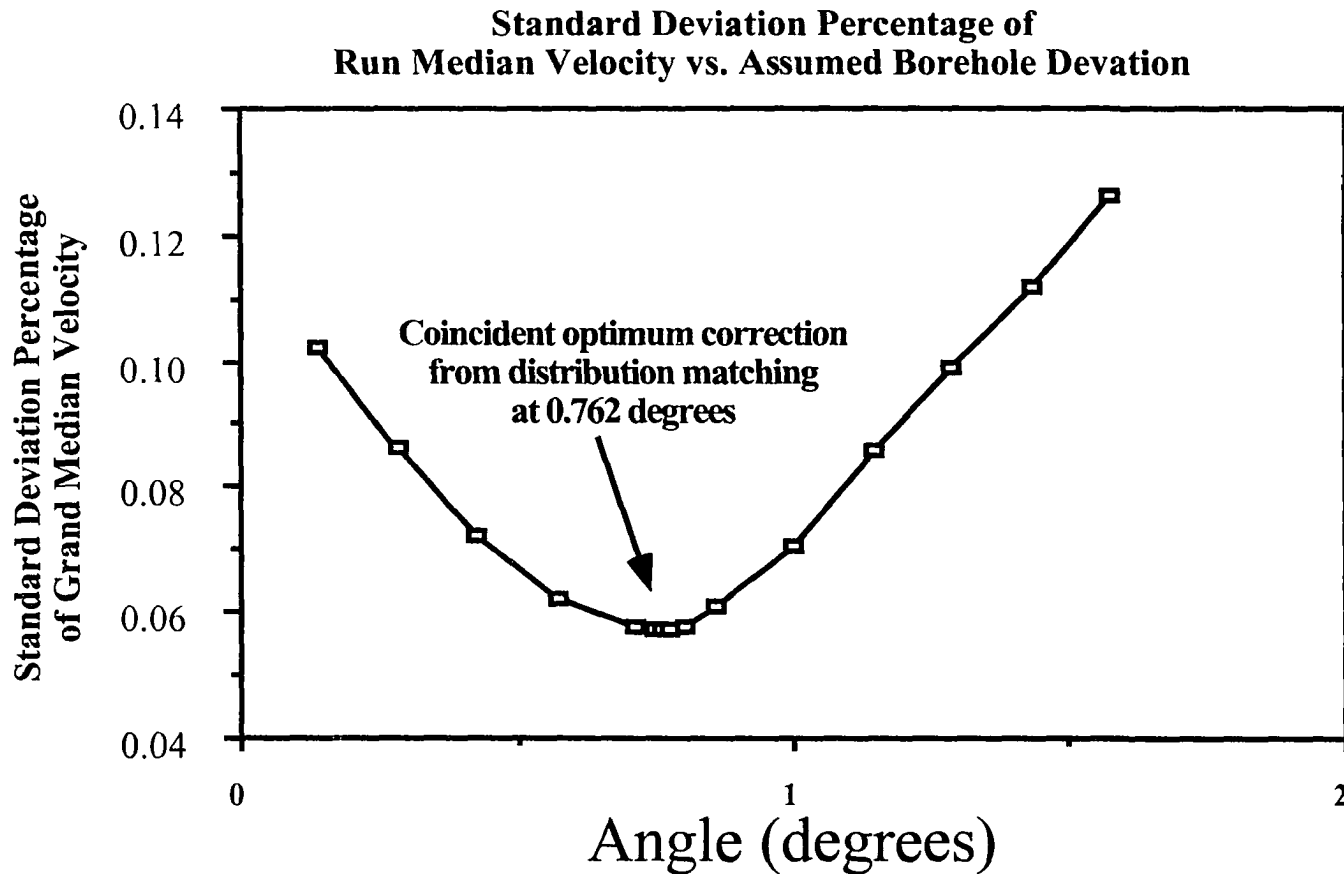


Figure 16: Stability of borehole deviation correction estimates indicated by a normalized dispersion measure. The standard deviation of median apparent velocity between sampling angle partitions is calculated as a function of assumed borehole deviation. It is then normalized relative to the grand median velocity. A well defined solution is qualitatively indicated by a distinct minimum at the point selected by distribution matching optimization.

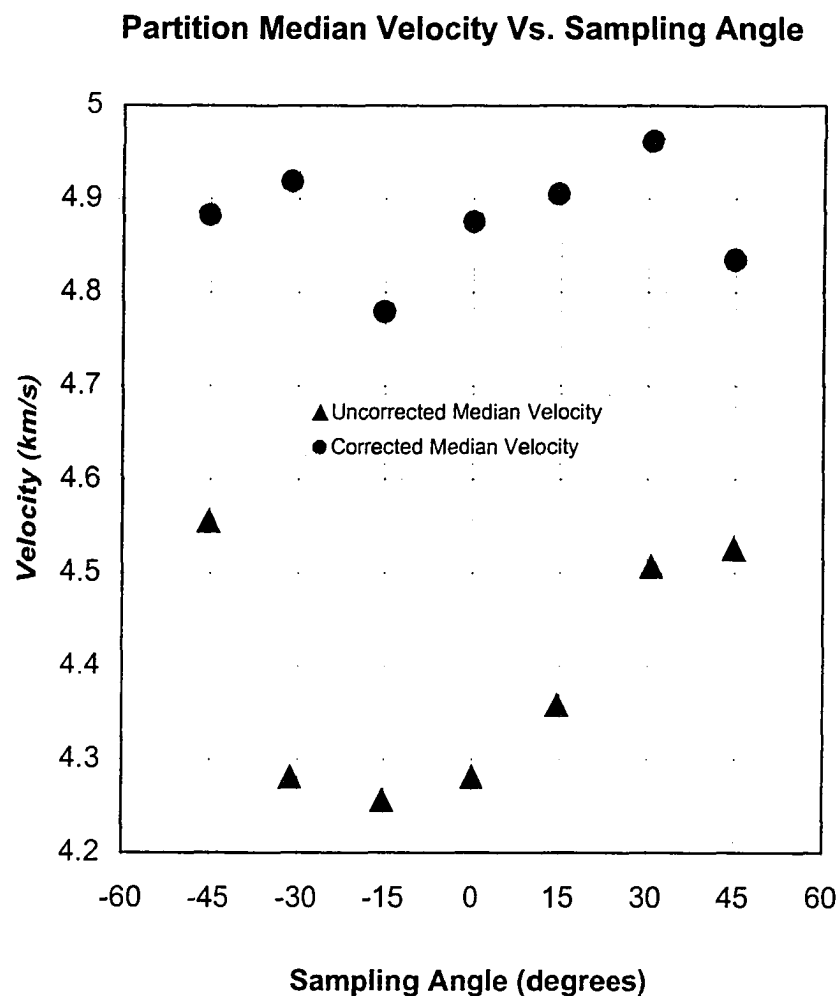


Figure 17: Median velocity as a function of sampling angle.

The slower velocities (triangle) are computed using a vertical borehole assumption. The systematic variation of velocity as a function of angle arises as a consequence of borehole deviation. The faster velocities (filled circles) are computed after optimizing the borehole deviation estimate using empirical velocity distribution matching.

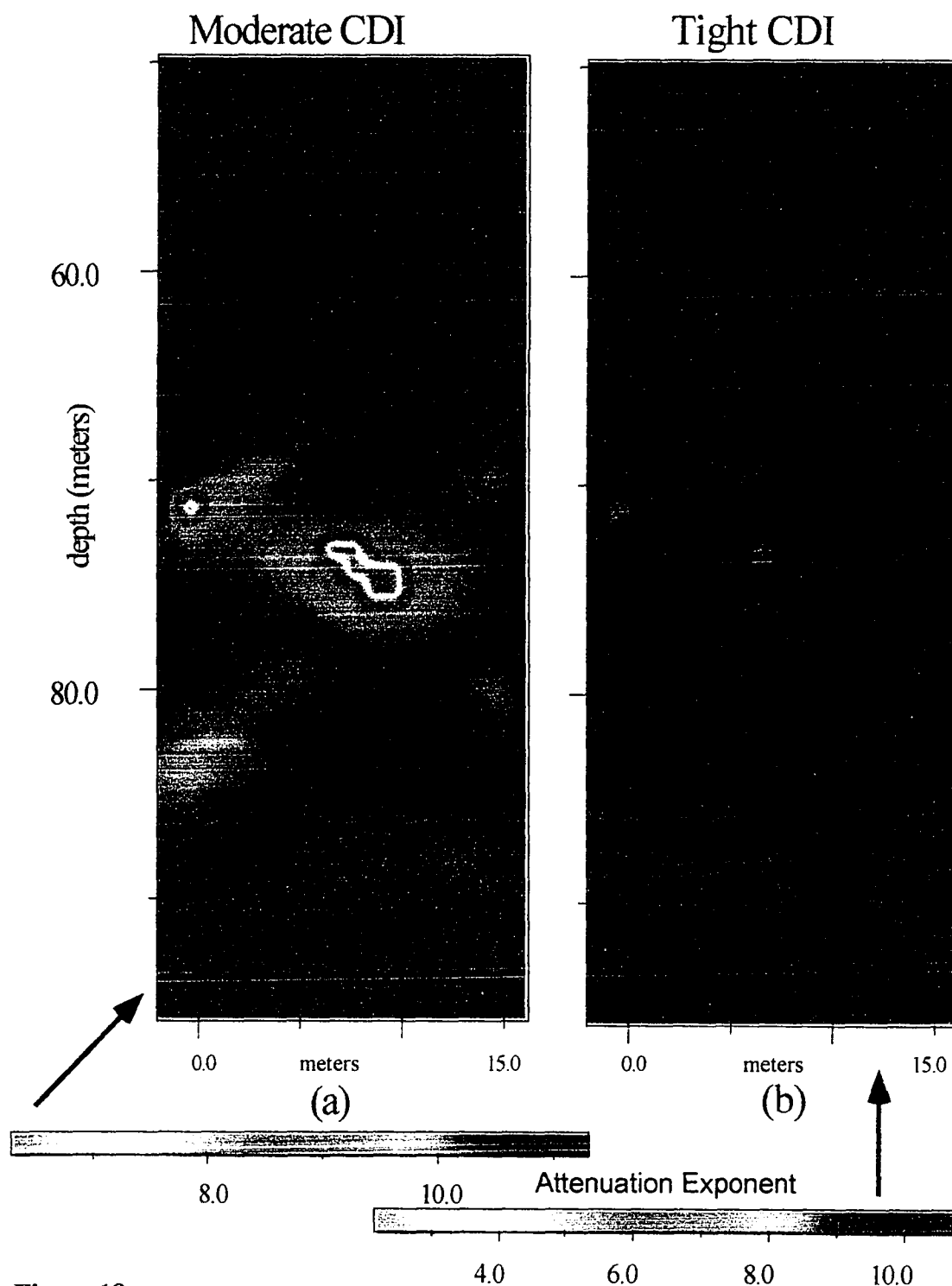
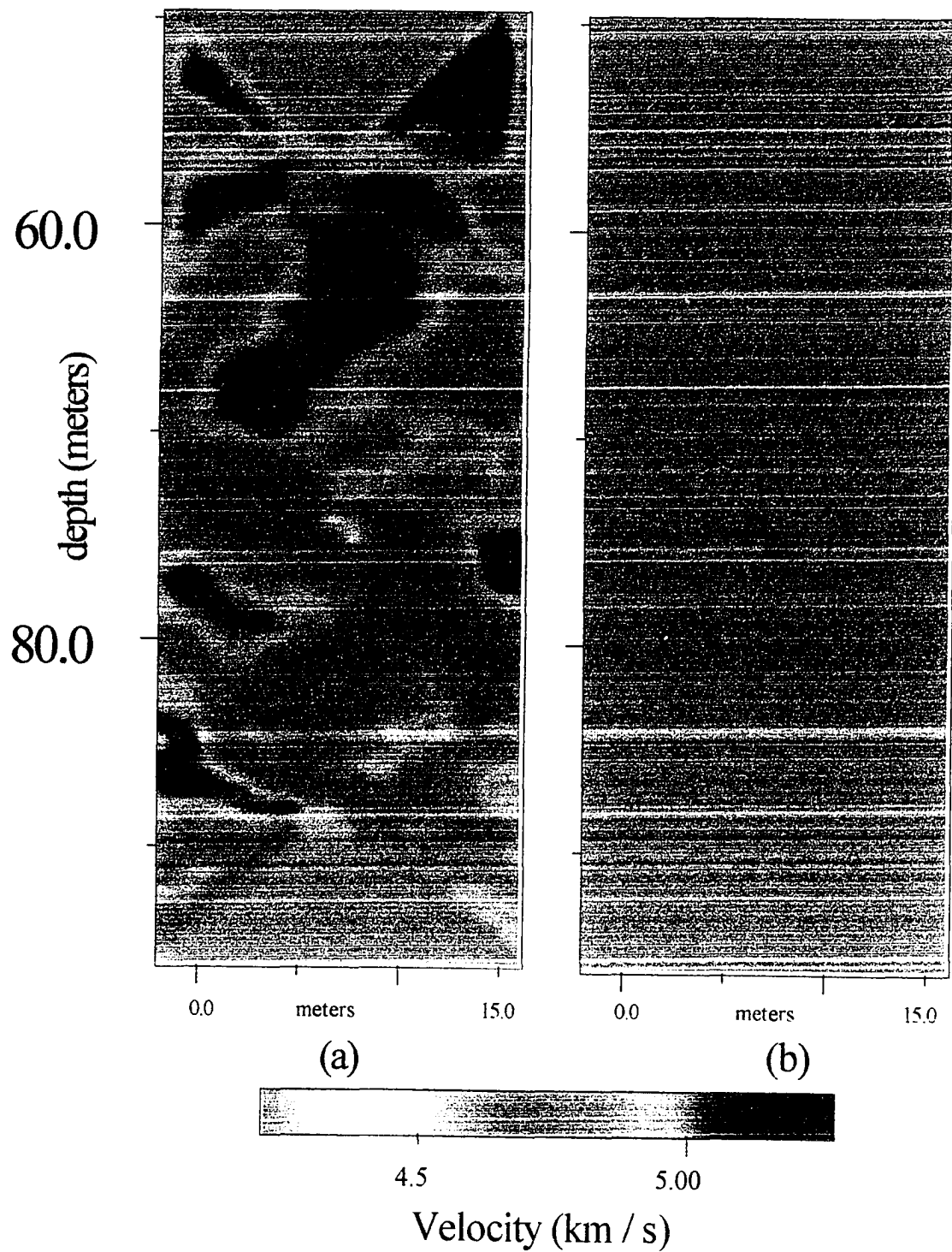


Figure 18: Final CDI attenuation tomograms.

(a) Loose tomogram: $r = 0.5$, $\sigma_b = 2$, $\sigma_a = 3$, rms err = 0.052.

(b) Tight tomogram: $r = 0.75$, $\sigma_b = 2$, $\sigma_a = 4$, rms err = 0.062.



CHAPTER 4

3D Seismic Raytracing Using Graph Theoretic Traveltime

Fields and Ray Refinement

ABSTRACT

An efficient, accurate, and robust raytracing methodology based on pre-calculation of an approximate traveltime field, followed by approximate raypath recovery and refinement, is introduced. A primary advantage of the method is that traveltime fields need not be recomputed unless large changes to the underlying velocity models are made. Graph-theory based traveltime field calculation is used. The most efficient scheme suggested to date for managing expansion of the minimum time surface is presented. Raypath refinement is performed using a pseudo-bending variant, which is modified for greater efficiency, improved accuracy, and use with general models. Examples and comparisons with previous work using both analytic and grid-based models are shown. The best performance for high accuracy applications is obtained when the most sparse, yet approximately valid, initial raypath is used. The rate of ray recovery over scales typical of exploration objectives is shown to be in excess of 1000 rays/second with errors on the order of 0.01%, using a 200 MHz PC. Greater rates of raypath recovery are possible on higher performance platforms using the same ANSI C++ (Stroustrup and Ellis, 1994) implementation.

INTRODUCTION

Motivation and General Overview

Minimum time seismic raytracing and first-arrival traveltimes field computation are key components of many geophysical tasks. Migration, ray modeling, tomography, and hypocenter location all require accurate traveltimes and/or raypath calculation between specified points in a model space. Computation of rays and traveltimes to required accuracy can be the most time consuming component of a processing task.

Computational efficiency, or the lack thereof, can mean the difference between an interactive and a batch process. Efficient traveltimes and raypath calculation translates directly into enhanced ability to test hypothetical models and to achieve higher accuracy results. This has motivated a great deal of work in these areas, and many methods have been proposed in the quest for a viable, general, solution.

Two basic approaches to two-point traveltimes computation through general models are possible: 1) Explicit ray shooting or bending methods (Julian and Gubbins, 1977); and 2) Traveltimes field computation methods, which may be further subdivided into finite-difference (Vidale, 1988) and graph theoretic approaches (Nakanishi and Yamaguchi, 1986). Many explicit ray methods efficiently and accurately compute raypaths and traveltimes for heterogeneous models (Cerveny and Psencik, 1983; Keller and Perozzi, 1983; Langan et al., 1985; Cerveny, 1985; Um and Thurber, 1987; Hanyga., 1988; Virieux et al., 1988; Grechka and McMechan, 1996), given specific knowledge of a valid approximate raypath or shooting angle. All of these methods are designed to

function within a limited range of models, but are extremely accurate and efficient when used appropriately. Many more investigations have focused on performance enhancement of explicit ray methods (Keller and Perozzi, 1983; Muller, 1984; Chapman, 1985; Lyche and Morken, 1986; Pereyra, 1988; Sambridge and Kennett, 1990; Stone and Forbes, 1990; Virieux, 1991). Nevertheless, explicit ray methods remain prone to slow convergence and/or convergence to local minima in the absence of valid approximate initial parameters (Fischer and Lees, 1992; Cao and Greenhalgh, 1993). The critical component allowing reliable and efficient ray shooting or bending is the ability to rapidly estimate good initial shooting angles or raypaths in general media. We believe that the most efficient method of estimating valid approximate raypaths is to recover them from a pre-computed approximate traveltimes field.

Traveltimes field computation methods are usually based on the computation of traveltimes at a set of grid nodes distributed throughout a target volume (Klimes and Kvasnicka, 1994; Vidale, 1988). Approximate raypath estimation based on path recovery from pre-computed fields requires negligible time relative to any explicit ray method. The accuracy of traveltimes field computation methods gradually increases as the grid-node spacing is decreased. Computation time and memory demands increase rapidly with decreasing node spacing, however. The net result is a class of methods that allow very efficient and robust estimation of approximate traveltimes fields and raypaths, but that can become computationally very expensive if high levels of accuracy are required.

Calculation of approximate traveltimes fields using graph-theory allows efficient and robust recovery of approximate raypaths through general heterogeneous media (Moser,

1992). Although the accuracy of the approximate raypaths and traveltimes may not be adequate for many applications, refinement of approximate raypaths to required accuracy levels is possible. We employ a variation of pseudo-bending (Um and Thurber, 1987), modified for a greater range of applicability and greater efficiency, to refine the approximate raypaths. A key advantage of this approach is that very accurate raypaths can be computed without recalculation of traveltime fields after moderate velocity model revision. As long as valid initial paths can be recovered from the unperturbed model, the bending algorithm rapidly converges to the desired raypath through the perturbed model.

The tradeoffs between traveltime field computation time, ray refinement time, and accuracy needs can be optimized based on the number of raypaths required, and empirical measures of performance for a given velocity model. Note that if refinement is always used, the traveltimes themselves need not be stored. It is sufficient to only store the spanning tree (Moser, 1993) for the target volume so that the approximate raypath may be recovered from any point. The ray refinement algorithm subsequently assigns traveltimes.

In summary, each single-strategy raytracing method performs optimally for some specified class of velocity models and some limited accuracy level. The objective of this paper is to describe a general and efficient hybrid approach to raytracing using algorithms appropriate for the accuracy level implicit at each stage of the problem. The hybrid approach efficiently provides solutions to the two-point raytracing problem through arbitrarily heterogeneous media. Graph-theory is used to provide approximate raypaths and a variation of pseudo-bending is used for ray refinement.

Several algorithms for computing graph-theoretic fields existed prior to this work (Nakanishi and Yamaguchi, 1986; Moser, 1991; Fischer and Lees, 1993; Cao and Greenhalgh, 1993; Klimes and Kvasnicka, 1994; Cheng and House, 1996). The basic pseudo-bending algorithm also existed (Um and Thruver, 1987). The primary contributions of this paper are:

- 1) Introduction of the most efficient method of graph-theoretic traveltime field computation described thus far.
- 2) A comparison of performance and errors associated with various implementations of graph-theoretic traveltime field calculation.
- 3) Extension of the pseudo-bending algorithm to achieve greater efficiency, accuracy, and robustness for use with general models.
- 4) Development and evaluation of an integrated raytracing method using graph-theoretic traveltime fields in conjunction with modified pseudo-bending ray refinement.

In the first portion of the paper, the rationale for using graph-theory based methods rather than the finite-difference approach is briefly discussed. Then a brief review of graph-theoretic traveltime calculation is provided. Factors affecting performance of graph-theory based raytracing are presented, and the most efficient possible queueing structure is introduced. In the second portion of the paper, a modified version of the pseudo-bending algorithm (Um and Thurber, 1987) is presented. We find that modified pseudo-bending is efficient and reliable for use with general velocity models given a valid initial raypath. Numerical investigations of computational efficiency and accuracy

are reported in the third portion of this paper. All examples are limited to isotropic velocity models, though modification of the algorithms for anisotropic models is possible. Extension of the proposed methods to maximum energy raypath estimation is beyond the scope of this paper.

Design Philosophy

Inadequacies with both explicit-ray and traveltimes computation methods currently used are attributable primarily to the algorithmic procedures and data representations used in implementation. All methods that we have investigated are based on sound principles of physics, and the mathematical approximations used are acceptable. For example, traveltimes field extrapolation using the finite-difference approximation of the eikonal equation (Qin et al., 1992) is fundamentally acceptable, but the algorithm most commonly used for implementation violates causality by extrapolating times along the surface of a cube, rather than along a minimum time surface (Vidale, 1990). Global minimum-time raytracing using graph theory is a valid concept (Nakanishi and Yamaguchi, 1986), but the requirement that raypaths pass through a finite set of grid nodes limits accuracy if reasonable computational resources are used. The basic pseudo-bending procedure seeks global minimum time paths by sequentially minimizing local segments, which is valid, but relies on the existence of velocity gradients. Any smooth raypath parameterization is only applicable within smooth media. Use of smooth raypath parameterizations in discontinuous models invariably ties whatever raytracing method is used to a fixed model representation. For each approach, limitations arise as a direct

consequence of the algorithmic implementation or fixed data representation. The components of a general solution seem to exist, but all components fail due to algorithmic or data representation limitations when pushed toward general application.

In light of the above observations, the problem of formulating a general method of efficient raytracing is primarily one of algorithm architecture. The design goals of true generality, data/method interchange, 'fair' algorithm comparison capability, and application extendability, can be achieved through an abstracted interface framework (Stroustrup, 1991). An abstracted framework describes conceptual objectives and data models, without reference to any specific representation (Buzzi-Ferraris, 1993). We adopt the philosophy of abstract design to achieve an implementation that can utilize, test, and compare the most appropriate existing methods within the appropriate contexts. Methods and representations can then be substituted, or used in conjunction with each other, to achieve a more universally applicable methodology than any one procedure used independently. When proposed methods seem viable for some portion of the raytracing task, they can be optimized, extended, and compared with methods currently in place without destabilization of the existing application. This results in a research platform capable of evolving toward higher levels of performance and a greater range of applicability. The remainder of this paper focuses on the most general, highest performance integrated algorithm methodology that we have tested within our abstract application framework.

TRAVELTIME FIELD ESTIMATION

General Algorithm Selection

Finite-difference traveltime calculation, as originally proposed, is based on traveltime extrapolation using the eikonal equation (Vidale, 1988). The number of available raypaths from a given source to any point in space is infinite, and local raypath curvature between nodes of the grid is considered. Graph-theoretic (also referred to as network-based), traveltime field computation (Cheng and House, 1996) is based on the minimum time transportation problem of operations research (Dijkstra, 1959). Raypaths travel only between grid nodes, local raypath curvature is ignored, and a finite number of available raypaths exist between any two-points (Moser, 1991). Finite-difference methods seem intuitively preferable to graph theoretic approaches, though it has been previously reported that finite-difference methods are no more accurate than graph-theory based methods (Klimes and Kvasnicka, 1994). Selection of one method over the other requires careful consideration of the ultimate application, accuracy, and robustness required.

Finite difference traveltime field computation (Vidale, 1988, 1990) is based on the extrapolation of traveltimes along the faces of an expanding cube (rectangle in two-dimensions). The method only requires modest computational resources and can be very accurate. The fundamental objection to the algorithm is that causality is obviously violated, since wavefronts expand along minimum time surfaces, not along geometric boundaries. Qin et al., (1992) demonstrated that traveltimes estimated using the finite-difference approach can be inaccurate in the presence of large velocity contrasts. In fact, large velocity contrasts aren't required to induce algorithmic failure. Any model that

implies the existence of turning, refracted, or reflected first arrivals, is potentially problematic. This has been recognized, and various modifications and reformulations of the basic finite-difference algorithm have been proposed (Van Trier and Symes, 1991; Liu, 1991; Watrus, 1991; Podvin and Lecomte, 1991; Pusey and Vidale, 1991; Schneider et al., 1992; Qin et al., 1992; Qin and Schuster, 1993). These algorithmic extensions provide more robust performance, but require more computing time than the original algorithm. Sufficiently accurate results are still not guaranteed for general media.

Another drawback to the finite-difference method occurs when partial derivatives of traveltimes with respect to model parameters are needed (for example, in tomography). The traveltimes at each node are calculated using linearized extrapolation, and it is prohibitive to save all partial derivatives of traveltimes in a manner that is consistent with the actual equations used. In practice, discrete raypaths are recovered by traveling down the traveltime gradient from a target point to the source location assumed during traveltime computation (Hole, 1992). Partial derivatives of traveltimes with respect to the model parameters are then computed along the raypath, but these derivatives and traveltimes are not consistent with the traveltimes computed using the extrapolation equation. The consequences of this inconsistency may be material, depending upon the model parameterization, grid spacing, and accuracy needed.

Graph theoretic methods are based on the expansion of a minimum time surface, and hence observe causality. Graph-theoretic methods robustly and efficiently compute approximate first-arrival traveltimes through models of arbitrary complexity. The traveltimes are approximate, however, because raypaths are required to travel only

between grid-nodes, and local raypath curvature is ignored. Theoretical analysis of graph-theory raytracing algorithm accuracy has been previously investigated (Moser, 1992; Klimes and Kvasnicka, 1994). Computational experiments indicate marginal accuracy improvements require a rapidly increasing amount of computer memory and CPU time as the last few percent of error is removed. This means that while graph theory based traveltime field computation robustly and efficiently produces traveltimes that are accurate to within a few percent, very large relative computation times are required to achieve high accuracy levels.

In summary, finite-difference methods can produce high levels of accuracy using reasonable computational resources, but are prone to failure as a consequence of causality violation. Graph theory based methods are extremely robust, but are only feasible if accuracy of no better than a few percent is required. Our strategy is to only use the traveltime field for initial raypath recovery. We use the graph theory based method to realize maximum robustness and efficiency at the accuracy levels required.

Graph Theory Traveltime Field Calculation

The objective of this section is to describe the basic graph-theory algorithm. Modifications introduced in subsequent sections are described within the context of the basic algorithm described here. The basic algorithm closely follows the structure of Klimes and Kvasnicka (1994) and their set nomenclature is retained. Both the breakdown of specific algorithm steps and cost analyses are slightly modified from the presentation of Klimes and Kvasnicka (1994). The modifications are made to allow a

more complete analysis of efficiency and to permit description of the algorithm within the context of an object-oriented design paradigm..

Graph theoretic raytracing (Nakanishi and Yamaguchi, 1986) is based on the principles of Fermat and Huygen (Aki and Richards, 1980). The solution is achieved using a method developed for the transportation problem of operations research (Dijkstra 1959). The transportation problem requires finding the minimum cost path from some node in a network to all other nodes in the network. A set of possible connections between the nodes, and associated costs for travel along the connections (*weights*), is considered known or easily computed. For raytracing, the cost of travel between nodes is the time required for a seismic wave to propagate between the nodes. The minimum traveltime path can be found using the Dijkstra algorithm applied to a mixed system of regularly space grid nodes and irregularly spaced source or receiver nodes. Raypath segments are allowed between any node in the network, and a finite number of nearby nodes. The collection of vectors, or *edges*, defining the possible connections between any node and all other nodes in a finite neighborhood is called the *forward-star*. The *weight* of each edge of the forward-star is defined by the traveltime between the node at which the forward-star is centered, and the node to which the edge indicates a connection. Specification of a prototype forward-star is required before computing the traveltime field. This includes specification of the maximum edge-length, h_{\max} . The following sets are then defined:

\mathcal{S} : A set of sources with known location and source-time;

\mathcal{N} : The set of all nodes in the network;

\mathcal{A} : The nodes at which minimum traveltimes are known;

\mathcal{B} : A *fringe* of nodes connected to at least one node of \mathcal{A} , but not members of \mathcal{A} ;

\mathcal{C} : Nodes of \mathcal{N} which are neither members of \mathcal{A} , nor \mathcal{B} .

Fringe nodes (members of \mathcal{B}) are connected to the set of nodes with known traveltimes (members of \mathcal{A}) via at least one edge of a forward-star extending from the known node to the fringe node. Candidate traveltimes for each fringe node are computed when the first connection to the node is made. The traveltime field expands by selecting the fringe (candidate) node with the minimum traveltime, and declaring it *final*. When a node is declared final all nodes connected to it must *updated*. Updating consists of revising candidate traveltimes downward if a potentially smaller time arises as a consequence of the newly declared final node.

An object-oriented (Buzzi-Ferraris, 1993) implementation of the algorithm relies on the following basic abstract *classes* (Stroustrup, 1991):

- 1) Node information representing locations connected to the network (Cnode).
- 2) A regular grid of Cnodes (Cgrid).
- 3) A descendent of Cnode representing source locations (Csource). Source locations are not necessarily coincident with locations on the grid. Sources have a known maximum time at which some arrival is observed. This time is usually used to specify the source time, but may also be used to specify a known time of arrival at a point. The union of all sources and all nodes of the grid constitute the set \mathcal{N} .
- 4) A queue of node pointers to all nodes $\in \mathcal{B}$ (Cqueue).
- 5) A class providing connection edges for any source or node (Cforward_star).

Typical data members of Cnode include:

- a) travel_time = traveltime observed at the node, null while node $\in \mathcal{A}$.
- b) queue_index = the position of the node in the queue, if the node $\in \mathcal{B}$. This data member is required if a sorted queue is maintained, and entirely superfluous if an unsorted queue is used.
- c) vs = velocity or slowness at the node;
- d) is_final = a flag indicating whether or not the node $\in \mathcal{A}$;
- e) back_index = an index of the previous node along the raypath;

The optional fields (c-e) enhance speed, but obviously increase the amount of memory required. The back_index data member is only useful if approximate raypaths are to be recovered directly from the grid connections. Using these classes, the traveltime field computation algorithm may be implemented as follows:

- 1) Initialization:
 - a) Set all nodes in the Cgrid to members of \mathcal{C} . This is accomplished by clearing is_final and setting traveltimes to null.
 - b) Move inaccessible nodes, such as those above the topographic surface to \mathcal{A} . This is accomplished by setting is_final = true, and travel_time = ∞ .
 - c) Enqueue all source nodes.
- 2) While the queue is not empty:
 - a) Dequeue the next node and declare it final, i.e., move it from \mathcal{B} to \mathcal{A} .
 - b) For each node, N_i , directly connected to the dequeued node:

If ($N_i \notin \mathcal{A}$)

Compute a new possible time for the node, τ_{new} .

If ($N_i \in \mathcal{C}$)

i) Move N_i from \mathcal{C} to \mathcal{B} , with time τ_{new} (i.e. enqueue the node).

Else if (τ_{new} is less than the existing candidate traveltime at N_i)

ii) Set the node traveltime to τ_{new} , and adjust the node position in the queue (or requeue the node), if required.

The cost of the above algorithm is approximately given by

$$C \cong n m f_i k_i + n k_{q+}(m, n) + n k_{q-}(m, n) + n m f_r k_{qr}(m, n) \quad (27)$$

where

n = total number of grid nodes requiring traveltime calculation

m = number of edges in a forward - star

f_i = mean fraction of forward star edges requiring traveltime calculation

f_r = mean fraction of forward star edges requiring a requeue, step (2.b.ii)

$k_i(m, n)$ = cost of calculating the traveltime between two nodes

$k_{q+}(m, n)$ = cost of enqueueing a node, step (2.b.i)

$k_{q-}(m, n)$ = cost of dequeueing a node, step (2.a)

$k_{qr}(m, n)$ = cost of requeueing or a node queue, step (2.b.ii).

The first term is the cost of actually computing traveltimes between nodes. It is a linear function of both the number of grid nodes used, and the number of edges in the forward-star selected. The second and third terms are the costs associated with queueing and dequeueing each node. Various possible queues incur different costs as a function of the

number of nodes in the queue, which is dependent on both n and m . The fourth term reflects the cost of moving, or requeueing, a node. This is required in step (2.b.ii).

GRAPH THEORY ALGORITHM OPTIMIZATION

Given error criteria and target volume dimensions, the minimum number of nodes in the network is fixed. The mean fraction factors of equation (27) are also fixed, given a grid, a forward-star, and a velocity model. This means that performance improvements can only be achieved by addressing one of three factors: 1) The cost of calculating a traveltime between nodes; 2) The number of edges and structure of the forward star; and 3) The cost of queue maintenance. A number of investigations have proposed efficiency improvements based on forward-star modification (Fischer and Lees, 1993; Cao and Greenhalgh, 1993; Klimes and Kvasnicka, 1994). Several investigations have also focused on queue efficiency (Moser, 1991; Klimes and Kvasnicka, 1994; Cheng and House 1996). In this section, the structure and possible methods associated with the cost of calculating a traveltime for a single edge is first reviewed. Next, the logic for adopting the forward-stars proposed by Klimes and Kvasnicka (1994) is discussed. The section concludes with the development of the most efficient queue structure possible.

Ray Segment Traveltime Calculation

The cost of calculating a traveltime along an edge of the forward star is optimal, given a traveltime equation, if recomputed quantities are eliminated and access time to required variables is minimized. This means that the slowness of each node should be

stored with the node, and the length of each edge of the forward-star should be stored within the Cforward_star class. The fastest method for calculating the traveltime between node i and node j is computed using the equation

$$\tau_{ij} = 0.5 d_{ij} (c_i^{-1} + c_j^{-1}) \quad (28)$$

where

τ_{ij} = traveltime between nodes i and j

d_{ij} = Euclidean distance between nodes i and j

c_i^{-1} = slowness at node i .

Other equations could be used to marginally improve the accuracy of estimated traveltimes, at the expense of computation time. Equation (28) assumes that the slowness field varies linearly along a straight path between the two nodes. If the velocity field is assumed to vary linearly between points, the traveltime equation is given by

$$\tau_{ij} = \frac{s_{ij}}{c_j - c_i} \log\left(\frac{c_j}{c_i}\right) \quad (29)$$

where s_{ij} is the distance between nodes i and j .

Equation (29), when protected from the case of small $|c_j - c_i|$, never predicts a traveltime that is larger than that computed using equation (28) and results in more accurate estimates of the traveltime field. We have found that the improvement can be noticeable for some velocity fields and grid-node spacings. Other possibilities exist. If a velocity field with continuous gradient information is available, analytic traveltimes between nodes using raypaths associated with circular arcs is possible. The finite-difference equations (Vidale, 1990) might also be used, when sufficient adjacent nodes have

traveltime information available. Since we primarily use traveltime fields only to recover approximate raypaths that are subsequently refined, differences in the node-to-node traveltime calculation are only material if the raypath is significantly altered by the equation. Computational experience indicates that this is not the case, and we use the least computationally expensive method, equation (28).

Efficient Forward-Star Design Details

The number of forward-star edges requiring traveltime calculation in algorithm step (2.b) largely controls the overall computation time required, given an efficient queuing method. Accuracy is largely controlled by the maximum gap between edges of the forward-star given a sufficiently small node spacing, hence forward stars that span many node intervals produce more accurate traveltime estimates. Early implementations of graph-theoretic raytracing employed cubical forward-stars (Moser, 1991; Saito, 1989). Spherical forward-stars contain far fewer edges for any angular gap threshold than cubical forward-stars (Klimes and Kvasnicka, 1994), and are always preferable. Other efficient forward-stars may be designed for specific representations of the velocity (Fischer and Lees, 1993). Spherical forward stars are the most efficient for use without reference to a specific velocity model representation, and are used throughout the remainder of this paper.

The radius of a spherical forward-star should be selected with knowledge of both the maximum possible curvature of the velocity field and some maximum acceptable error threshold. For a rectangular grid, the angular gap is a piecewise constant function of the

forward-star radius. Optimal forward stars are designed (Klimes and Kvasnicka, 1994) according to the criteria:

$$h_{\max} = h_{\min} \sqrt{n_{fs}^2 + (d_p - 1)} \quad (30)$$

where

h_{\max} = maximum edge length (radius of the forward - star)

h_{\min} = grid node spacing

n_{fs} = the 'level' of the forward star = the half - length
of the smallest cube containing the
forward - star, expressed in grid intervals

d_p = dimensionality of the problem, i.e. 2 for 2 - D, 3 for 3 - D.

Forward-stars with h_{\max} other than that given by (30) include an unnecessarily large number of edges with no reduction in maximum angular gap. The largest forward-star radius consistent with equation (30) that does not alias the velocity field produces the most accurate traveltimes in the most efficient manner.

The number of nodes in a forward star of any radius can be reduced without sacrificing accuracy by eliminating long edges within a small angular distance of any shorter edge (Klimes and Kvasnicka, 1994; Cao and Greenhalgh, 1993). The number of edges eliminated in the forward-star reduction process can be considerable (Figure 20 and Figure 21).

It has also been proposed that only a small subset of forward-star edges need actually be evaluated at any node (Cao and Greenhalgh, 1993). This follows from the fact that a local change in raypath direction must obey the eikonal equation in continuous media, and Snell's Law in discontinuous media. Ray-shooting methods are based on this

principle (Julian and Gubbins, 1977; Langan et al., 1985; Cervený, 1987). Since the previous connecting node is always known at any point in the traveltime field expansion process, it should be sufficient to evaluate a small number of edges in the vicinity of the direction predicted by shooting theory. Algorithms based on this idea have been termed 'dynamic directional-graph' methods, because a subset of forward-star edges are dynamically selected at each node (Cao and Greenhalgh, 1993). Though some robustness may be lost, this approach is observed to significantly improve efficiency when used with two-dimensional grids and edges sorted by angle, rather than length. For three-dimensional problems, no efficient organization of edges has been introduced that allows effective use of this concept.

In summary, we start with spherical forward-stars designed in accordance with equation (30). Long-edges of the forward-star within a small angular distance of any shorter edge are then removed. Edge lengths are stored with the forward-star, and the edges are applied from shortest to longest. A cross-section of a reduced, spherical forward-star prototype in the x-y plane illustrates the edges actually available for use (Figure 20).

Efficient Queuing

The original Dijkstra algorithm uses an unsorted queue. Enqueueing only requires appending an element to the end of an array, and computing expense is negligible. Dequeueing costs are large, however, because the entire queue must be searched for the minimum element every time an element is dequeued. The cost of dequeueing a single

node from a queue of length n_q may be expressed as $k_{q-} = k_c n_q$, where k_c is the cost of retrieving and comparing the traveltime of an element in the queue with another value. The gap left by the dequeued element is replaced by the last element in the queue as part of the dequeuing process. This original Dijkstra algorithm produces the desired results and is simple to implement, but computing costs are prohibitive for use with large queues.

Computing costs can be improved by efficiently maintaining a queue from which the minimum element may be retrieved with little overhead. A heap queue was used in the first efficient implementation of graph-theory based traveltime field computation (Moser, 1991). A heap is a binary-tree type of data structure, with elements, x_i , satisfying

$$x_i < x_{2i} \quad \text{and} \quad x_i < x_{2i+1}, \quad (\text{Knuth, 1968}). \quad (31)$$

Enqueueing consists of appending an element to the tail of the heap, and then performing comparison and swap operations up the binary path from the bottom until the value of the new element either satisfies equation (31) or reaches the top of the heap. Efficient dequeueing begins with removal of the element at the tail, and storing it for placement at a new location. The first element is then removed for use as the dequeued element. Then the stored element is placed in the first position, and comparison/swap operations are performed down the binary tree until the position of the relocated element satisfies equation (31). When using a heap, both queueing and dequeueing costs are

$$k_{q-,q+} \leq (k_c + k_s) \log_2(n_q), \quad \text{where } k_s \text{ is the cost of swapping elements in the heap.}$$

Requeueing operations are required any time a fringe node candidate traveltime is

reduced in algorithm step (2.b.ii). Costs are small because it is unlikely that the position of the node in the heap will require more than a few comparison/swap operations.

Other queue structures have also been investigated for use with graph-theory based raytracing. The LQueue algorithm is reported to be less efficient than the heap (Moser, 1991). A Quicksort queue is reported to be marginally faster than the heap (Cheng and House, 1996), but is still a process of order $\log_2(n_q)$. An unsorted queue used in conjunction with the BUCKET algorithm is reported to be much faster than a heap queue (Klimes and Kvasnicka, 1994), although our observations (described in the following paragraphs and in the results section) suggest otherwise.

The principle of the BUCKET algorithm requires a bounded velocity field, a fixed node spacing, and a known maximum forward star radius. In this case, there exists a minimum difference between the actual first arrival at any node, τ_{\min} , and the next possible arrival along any other path, $\tau_{\text{next}} > (\tau_{\min} + \delta\tau)$. It is therefore sufficient to select any node with traveltime less than $\tau_{\min} + \delta\tau$ from the fringe. This holds because paths through any node selected with a traveltime satisfying $\tau_{\text{node}} < \tau_{\min} + \delta\tau$ cannot possibly perturb the minimum time node (the minimum additional traveltime added to a node is larger than $\delta\tau$). The procedure is then to divide the queue into a set of intervals of size $\delta\tau$, starting from the source with the minimum time, and dequeue any element in the smallest non-empty interval. When using the BUCKET algorithm with unsorted queue, enqueueing cost is negligible, and is associated only with the process of appending an element to the end of an array. Requeueing is not required, and therefore costs are zero.

Dequeuing costs are associated with the time required to search for an element in the first non-empty interval, removal of the element, and filling the gap with the tail element. The number of operations required for dequeuing was proposed to be proportional to the square root of the number of edges in a forward star, $k_{q-} \sim \sqrt{m}$ (Klimes and Kvasnicka, 1994). We find that the number of operations required to search an unsorted queue for an element in the smallest non-empty interval is much greater than $\log_2(n_q)$ (Figure 22), and the algorithm is much less efficient than either the heap or quicksort approach. The comparative performance reported in Klimes and Kvasnicka (1994) used two different implementations. We suggest that the good relative performance reported is due to implementation differences other than queueing. The results reported here share a common, abstracted, interface: Cost differences only arise as a result of the type of queue interactively selected for *instantiation* (Stroustrup, 1991).

Although use of BUCKET with an unsorted queue is inefficient, a simple modification of the procedure can reduce queueing costs to a minimum level. Instead of using a single unsorted array, we use a queue representation consisting of an array of LIFO-stacks, one associated with each traveltime interval of size $\delta\tau$. Queueing operations, k_{q+} , consist of a floating-point division to locate the correct stack, and a push of the node pointer onto the stack. LIFO-stacks are one of the fastest possible structures for unsorted data storage and retrieval, and queueing cost can be considered minimal. Dequeueing an element requires moving the minimum non-empty queue pointer forward (if required), and popping an element off the stack. These are also a

negligible time operations. Stacks are instantiated as needed and are recycled from the front to the end of the queue as the early intervals are emptied. Note that elements must be requeued if a revision of traveltime in step (2.b.ii) does require requeueing. The requeueing process does not, however, require extraction of the element from its original stack, only another push onto the appropriate stack. When a previously finalized element is popped from a higher time interval stack, no operations other than the stack pop are performed. Requeueing does increase the memory required for the overall procedure, but has little effect on performance. The use of BUCKET with the LIFO stack queue results in a process with minimal queue overhead, and the time required for the overall process is approximately a linear function of mn .

PSEUDO-BENDING RAYPATH REFINEMENT

As previously indicated, and as will be shown in the results section, the inaccuracy associated with graph-theory based traveltime field computation is too large for many applications. Errors typically range from approximately one to five percent, using parameterizations that require reasonable quantities of memory and computation time. The associated raypaths are, however, fairly good approximations of the actual raypaths for any velocity structure. We therefore seek an efficient method to refine an approximate raypath to sufficient accuracy. Experiments with variations of shooting methods (Julian and Gubbins, 1977; Langan et al., 1985) proved to be disappointing. The initial shooting angle recovered from the approximate path is usually insufficiently accurate for rapid convergence. Complete failure is possible for several cases, the most

important of which are refracted arrivals. No shooting method that we know of properly handles the family of raypaths associated with a refractor in accordance with Huygen's principle. Error reduction of graph-theory based rays has been demonstrated using straight-path and Snell's law refinement (Fischer and Lees, 1993), but occasional failure was also reported. Refinement to arbitrary accuracy is not possible using Snell's Law or straight-path refinement in conjunction with gradational models.

Most explicit ray methods based on continuous models require reparameterization for use with discontinuous or block models (Lees and Shalev, 1992; Fischer and Lees, 1993). Ray-bending methods that rely on smooth media or analytic raypath descriptions (e.g. Prothero et al., 1988; Grechka and McMechan, 1996) are unacceptable for general use in the presence of topography or sharp subsurface velocity boundaries. Efforts and suggestions directed towards generalization of such approaches invariably tie the raypath representation directly to a piece-wise continuous model representation (e.g., Gjoystdal et al., 1984; Cerveny, 1987; Farra et al., 1989). Any fixed velocity model representation results in limited possibilities for application extensions. A great deal of effort can be spent trying to force a desired physical model description into an inappropriate framework when compared with that required to simply introduce a new, appropriate model within an abstract framework. In light of this, the potential for abstraction, extension, and general application are paramount considerations when selecting methods for use with arbitrary models.

Raypaths described as a sequence of nodes easily accommodate discontinuities, including topographic surfaces, and can be used without reference to a specific velocity

model. A variety of methods are available to optimize the location of node-based raypaths. In keeping with the philosophy of general design, we seek an extendable algorithm for bending node-based raypaths into an optimal geometry. Relaxation provides such an approach. The solution of differential equations using relaxation relies on the idea that the relationship between adjacent physical properties, or field values, must obey a specified differential equation, or law. Relaxation seeks a solution by iteratively estimating new parameter values such that the required relationship between parameters is honored. The minimum time raytracing problem seeks a global minimum time raypath. A basic theorem from operations research states that any segment of a global minimum path represents the minimum time path between the segment endpoints. This implies that the global minimum time raypath may be found by allowing each segment of a valid initial raypath to 'relax' into minimum time configuration.

Pseudo-bending (Um and Thurber, 1987) is fundamentally a relaxation method. The algorithm acts sequentially upon segments of node-based raypaths, always descends from an initial path toward a faster one, and is relatively easily modified for use with general velocity models. The original algorithm requires the existence of velocity gradients at all points, though it is possible to modify the procedure for general use. A general, adaptive, finite-difference method has been previously investigated (Pereyra et al., 1980), and is a viable alternative to the adaptive methodology outlined in the following paragraphs.

Basic Pseudo-Bending

The pseudo-bending algorithm operates sequentially on nodes of an initial raypath to find a smaller time path along each sequence of three nodes. Given three nodes along a raypath, \mathbf{x}_{i-1} , \mathbf{x}_i , and \mathbf{x}_{i+1} , the perturbation task is to reposition \mathbf{x}_i in a way such that the total traveltime from \mathbf{x}_{i-1} to \mathbf{x}_{i+1} , and passing through \mathbf{x}_i , is a minimum. When velocity gradients are available, the optimal location for \mathbf{x}_i is predicted to be:

$$\mathbf{x}'_i = \mathbf{x}_{\text{mid}} + R_c \hat{\mathbf{n}} \quad (32)$$

where

$$\mathbf{x}_{\text{mid}} = \frac{(\mathbf{x}_{i-1} + \mathbf{x}_{i+1})}{2}$$

$$\mathbf{n} = \nabla V_{\text{mid}} - [\nabla V_{\text{mid}} \cdot \bar{\mathbf{x}}] \frac{\bar{\mathbf{x}}}{\bar{\mathbf{x}} \cdot \bar{\mathbf{x}}}$$

$$\hat{\mathbf{n}} = \frac{\mathbf{n}}{|\mathbf{n}|}$$

$$R_c = \frac{-(\sigma V_{\text{mid}} + 1)}{4 \sigma \hat{\mathbf{n}} \cdot \nabla V_{\text{mid}}} + \sqrt{\frac{(\sigma V_{\text{mid}} + 1)^2}{(4 \sigma \hat{\mathbf{n}} \cdot \nabla V_{\text{mid}})^2} + \frac{|\mathbf{x}_{i+1} - \mathbf{x}_{\text{mid}}|^2}{2 \sigma V_{\text{mid}}}}$$

$$\bar{\mathbf{x}} = \mathbf{x}_{i+1} - \mathbf{x}_{i-1}$$

$$\sigma = \frac{V_{i-1}^{-1} - V_{i+1}^{-1}}{2}$$

and velocities at the required locations are given by

$$V_{\text{mid}} = V(\mathbf{x}_{\text{mid}}), \quad V_{i-1} = V(\mathbf{x}_{i-1}), \quad V_{i+1} = V(\mathbf{x}_{i+1})$$

$$V_{\text{mid}} = V(\mathbf{x}_{\text{mid}}), \quad V_{i-1} = V(\mathbf{x}_{i-1}), \quad V_{i+1} = V(\mathbf{x}_{i+1})$$

(Um and Thurber, 1987). Equation (32) is analytically derived assuming:

- 1) Locally straight raypaths between points is a valid approximation;
- 2) The traveltime between points is computed using equation (28);

3) $\nabla V \cong \nabla V_{\text{mid}}$ in the region of perturbation (a Taylor series truncation is used).

There is a small inconsistency in the derivation: A constant velocity gradient is assumed, which implies that traveltimes are more appropriately calculated with equation (29) than with equation (28). This, combined with the fact that the points \mathbf{x}_{i-1} and \mathbf{x}_{i+1} also need to be perturbed towards the correct raypath, provides motivation for introducing an *ad hoc* perturbation exaggeration factor, F . In actual practice this exaggeration factor is used to predict the replacement point for \mathbf{x}_i as

$$\mathbf{x}_i'' = \mathbf{x}_i + F(\mathbf{x}_i' - \mathbf{x}_i). \quad (33)$$

The value of F that provides the best performance is a function of both the velocity model and the initial raypath inaccuracy. Usually F is chosen such that $(1 \leq F \leq 2)$. For typical models and initial raypaths provided by graph-theory, good performance is achieved by choosing F such that $(1.1 \leq F \leq 1.3)$ (Um and Thurber, 1987). Using equations (32) and (33) to predict node perturbations, the basic pseudo-bending process may be described by the pseudo-code below.

Pseudo-code for the Pseudo-Bending Algorithm

```

do
  do
    sequentially perturb all nodes
    while (traveltime reduction > p)
      double the number of nodes in the raypath
  while (traveltime reduction > p)

```

Accuracy is usually a factor of 5-10 times larger than the convergence threshold, p .

Performance Enhancement

Substantial performance enhancement of the pseudo-bending algorithm is possible by modifying a few subtle aspects of the procedure. The following modifications are made to achieve higher rates of convergence and more accurate raypaths/traveltimes:

The relaxation schedule is modified so that sequential perturbation of nodes begins at both ends of the raypath. The perturbation operators cross at the middle node, and continue until the other end of the raypath is reached. We refer to this as the “end to end” relaxation schedule. The original relaxation schedule suggested halting the perturbation operators when they collide at the middle node, which we refer to as the “stop at mid-node” schedule. This implies one-sided ‘attack’ directions, which results in decreased accuracy and slow convergence when large regions of small velocity gradient are encountered. With uniform gradients, the stop at middle node approach results in slightly faster convergence rates (Figure 23a), but with larger errors (Figure 23b). An “alternating end” relaxation schedule is also a possibility, which is implemented by starting a single perturbation operator at one end of the raypath, and reversing its direction when the other end of the raypath is encountered. This approach produces errors that are similar to the end to end approach, but at considerably more computational expense. A comparison of typical cost and accuracy for the three relaxation schedules and thirteen rays in an exponential model (Figure 24) is illustrated (Figure 25).

Ray segment traveltimes are computed assuming linear velocity variation, equation (28), rather the linear slowness variation relation, equation (29). This always results in

smaller traveltime estimates. For a fixed convergence threshold, memory requirements are usually reduced by 50% and computation times are typically reduced by 25-40% relative to that required when using the linear slowness variation equation. Typical performance improvements are illustrated in Table 1. During node doubling, the newly inserted nodes are assigned coordinates predicted by either equation (32) or (33), instead of the mean location. This is required if linear velocity variation is used to predict traveltimes, since the original procedure does not include traveltime reduction associated with bending to indicate overall convergence. Instead, the original algorithm uses the idea that the traveltime predicted by equation (28) asymptotically approaches that predicted by equation (29) as the node spacing becomes infinitely small. In first-order media, equation (29) always computes the same traveltime along any fixed path regardless of the node spacing used during computation. This means that if the original criterion were used in conjunction with equation (29), convergence would always be indicated, unless an aliased low velocity zone were found during node-doubling.

The pseudo-bending algorithm described above is reliable and efficient for sufficiently smooth models with no large constant-velocity regions. When failure occurs, it is generally because the analytic perturbation predicted by equation (32) is only valid for velocity gradients that are approximately constant over the region of perturbation. Analytic perturbations estimated in the vicinity of strong velocity field curvature are erroneous, and do not effectively minimize ray segment traveltimes. If accepted, these estimates can cause spurious node placement, slow convergence, and raypath inaccuracy.

We have actually encountered several cases of complete convergence failure when raypaths are perturbed near ridges of high velocity.

Velocity field curvature, at scales that cause problems for pseudo-bending, can be caused by discontinuities, large curvature of a continuous field relative to the current ray node separation, or the existence of a topographic surface. Failure associated with velocity field curvature may be addressed by verifying the linearity assumption before accepting a perturbation prediction. If velocity field curvature is detected, an alternate perturbation scheme must be used.

Strongly Non-Linear Models

Modification of pseudo-bending to handle general non-linearities in a continuous velocity field is fairly straightforward. The initial perturbation location, \mathbf{x}'_i , is first predicted using equation (32). The actual velocity at this location, $V(\mathbf{x}'_i)$, is then evaluated and compared with the velocity predicted by a first-order Taylor series approximation, $V_{\text{pre}}(\mathbf{x}'_i) = V_{\text{mid}} + R_c (\hat{\mathbf{n}} \cdot \nabla V_{\text{mid}})$. If the prediction error fraction is larger than a specified nonlinear detection threshold, $\left\{ \left| V(\mathbf{x}'_i) - V_{\text{pre}} \right| V(\mathbf{x}'_i)^{-1}_{\text{mid}} > \varepsilon \right\}$, then a region of large velocity curvature is declared, and the original perturbation scheme is abandoned in favor of a specialized approach. To achieve full protection against the deleterious effects of velocity field curvature, consistency of predicted velocity with observed velocity at the end-points should also be ensured. Practically this is not required if a good initial raypath is available, since the primary concern is elimination of

exaggerated perturbation offset estimates. Once an unacceptable non-linearity is detected, there are several options for handling the condition. The primary goal of the condition handlers is to eliminate large estimates that lead to convergence failure or unacceptable raypath accuracy resulting from large, incorrect node perturbations.

The most robust method of handling non-linearity is to directly search the plane passing through \mathbf{x}_{mid} and normal to $(\mathbf{x}_{i+1} - \mathbf{x}_{i-1})$ for the point in the plane producing the minimum three-point time. Nonlinear conjugate-gradient methods (Polak, 1971) can be used to find the optimal point for models with gradient information. The cost of using the method is approximately equal to the cost of two line-searches over a very small distance. If gradient information is not available, downhill simplex (Nelder and Mead, 1965) is used. Downhill simplex requires far more computational time than conjugate gradient, but is robust, and does not require gradient information. Both methods require many evaluations of the velocity model at specific points. This is relatively inexpensive for grid-based models, but can be prohibitively time consuming for series-based models.

Since the perturbation process is iterative, it is not necessary to use truly optimal perturbation estimates. Recall that the perturbations indicated by both equations (32) and (33) are not exact, and are only approximate steps in the right direction. They are used in the interest of overall performance, and iteration is relied upon to produce convergence to the desired raypath. Perturbation estimates made in response to detected nonlinear velocity variation may also be approximated. Both conjugate-gradient and downhill simplex produce very accurate estimates of the optimal perturbation, but they do so at considerable computational expense. Less accurate perturbations imply the need for

more iterations prior to convergence, but overall performance may be considerably improved if sufficiently simple and accurate approximations are used.

Simple perturbations designed to handle detected nonlinear conditions can be implemented using rules based on the direction of predicted velocity inaccuracy. If the actual velocity is greater than the predicted velocity, then the perturbation distance is underestimated, and it should be revised upwards. If the actual velocity is less than that predicted, the perturbation distance is overestimated, and the perturbation distance should be revised downward. For all approximate solutions, the direction of perturbation, $\hat{\mathbf{n}}$, is assumed to be approximately correct, but the distance estimate, R_c , is assumed to be inaccurate.

If $V_{\text{mid}} > V_{\text{pre}}(\mathbf{x}'_i)$, R_c is underestimated, its value should be increased. A single line-search directed outward from the predicted point along $\hat{\mathbf{n}}$ can be used, the estimate may be accepted as is, or a revision based on the observed mean gradient from \mathbf{x}_{mid} to \mathbf{x}'_i can be used. Experiments indicate that the line-search approach can be less efficient than use of conjugate-gradient, presumably because errors due to inaccurate estimates of $\hat{\mathbf{n}}$ are also addressed by the conjugate-gradient procedure. Errors due to inaccurate $\hat{\mathbf{n}}$ estimates are amplified by large R_c . Equation (32) indicates that large values of R_c occur only in the presence of large gradients. Perturbation search-plane optimization using the non-linear conjugate gradient procedure addresses these errors at a cost of only two to three times that required for a single line-search. If the estimate is accepted as is, the process falls into the restricted-step family of procedures. Restricted-step procedures are often used to limit potential perturbation overestimation in general non-linear optimization

(Fletcher, 1987; Bezier, 1972). Convergence continues safely, but the number of iterations required increases. Alternatively, a revised R_c can be derived using the observed mean gradient in the direction of $\hat{\mathbf{n}}$. Usually, this produces the best performance. Very occasionally, R_c revision results in slow convergence and raypath inaccuracy due to overestimation of perturbation distances. If R_c revision is found to generate unacceptable results for a particular model, restricted steps or direct perturbation plane optimization should be used.

If $V_{\text{mid}} < V_{\text{pre}}(\mathbf{x}'_i)$, then R_c is overestimated, and its value should be decreased. Either a single line-search on the line $(\mathbf{x}_{\text{mid}}, \mathbf{x}'_i)$, or a revision based on the observed mean gradient from \mathbf{x}_{mid} to \mathbf{x}'_i can be used. The initial perturbation distance cannot reliably be accepted because it is an overestimate. Observed non-convergence for this case motivated our original work to refine and extend pseudo-bending. Experiments indicate that the line-search approach usually performs adequately, and is faster than conjugate-gradient. This is presumably because the small R_c implicit with small gradients mitigates errors in $\hat{\mathbf{n}}$. Alternatively, a revised R_c can be derived using the observed mean gradient in the direction of $\hat{\mathbf{n}}$. On average, this produces acceptable performance, but very occasionally results in slow convergence and raypath inaccuracy due to incomplete compensation for velocity field curvature. For safe application, the velocity at the point predicted by the revised R_c should be iteratively examined for prediction error as R_c is revised downward.

Snell's Law, Constant Velocity Regions, and Topography

Discontinuities in a velocity model are a special case of a strong non-linearity, and are usually handled sufficiently well by the procedures described above. Occasionally, however, it may be desirable to ensure that Snell's Law is explicitly honored at an interface. The basic pseudo-bending algorithm does not honor Snell's Law, and modifications are required to accommodate this need. Given a velocity model representation, modification of the pseudo-bending algorithm to accommodate Snell's Law is straightforward. A method of applying Snell's Law in conjunction with basic pseudo-bending has been previously investigated (Zhao, 1990), but the method makes explicit reference to a specific velocity model representation. The problem is basically one of software architecture if the bending procedure itself is to be used with general models.

We seek a general, abstract design in which the interface between velocity models and the ray-bending algorithm is fixed for all models, and the representation hidden from the ray-bender. This facilitates introduction of new velocity model parameterizations as needed. In keeping with this principle, velocity models should not have access to the ray representation, either. To address these design goals, an interclass query interface is introduced, and specific helper classes are provided for each combination of velocity model and raypath representations.

All velocity model classes must provide a flag indicating whether or not explicit interfaces exist at which Snell's Law should be applied. For any class with such interfaces, the appropriate helper class can be provided to the ray-bender class. The

helper is instantiated during ray-bender initialization by a master helper manager. During bending, the ray-bender class queries all velocity models with the Snell's Law flag set for information as to whether a discontinuity exists between \mathbf{x}_{i-1} and \mathbf{x}_{i+1} , prior to perturbing node \mathbf{x}_i . If no discontinuity exists, node \mathbf{x}_i is perturbed in accordance with modified pseudo-bending. If a discontinuity does exist, the ray-bender invokes the helper class to modify both points \mathbf{x}_i and \mathbf{x}_{i+w} , given \mathbf{x}_{i-2w} and \mathbf{x}_{i-w} , such that Snell's Law is honored. Here, $w = 1$ for a node perturbation operator moving toward higher indices, and -1 for an operator moving from larger indices toward smaller ones. The node perturbation operator resumes operations at node \mathbf{x}_{i+2w} . This design allows flexibility for new velocity model representations, does not impose measurable efficiency penalties on continuous models, and does not sacrifice functionality relative to a single case implementation.

Constant velocity regions present a major obstacle for unmodified pseudo-bending. Regions of constant velocity predict no need for perturbation other than location averaging, because the velocity gradient is zero. This results in slow convergence and physically unacceptable raypath bending within the constant velocity region. When a constant velocity region is embedded between two regions with a non-zero gradient, a virtual information disconnect occurs. On both sides of the interface, the fastest path to the interface is selected with no regard for the traveltime through the constant velocity volume. The problem can be addressed by optimizing a four point traveltime problem with simultaneous perturbation of the two nodes representing the endpoints of the constant velocity region. Node placement is optimized within a search sphere, rather

than a plane. A sphere is required to avoid 'trapping' nodes against velocity boundaries, which results in restricted raypath motion and premature convergence detection.

Constant velocity regions are also handled by first detecting the region, whereupon alternate action may be taken. As a node perturbation operator advances along a raypath, a forward look is made to detect and mark the bounding indices of any sequence of nodes having identical velocities. If the first and last nodes in such a region are \mathbf{x}_i and \mathbf{x}_j , then the task is to simultaneously perturb both nodes \mathbf{x}_i and \mathbf{x}_j in such a way that the sum of traveltimes along the straight paths $(\mathbf{x}_{i-1}, \mathbf{x}_i)$, $(\mathbf{x}_i, \mathbf{x}_j)$, and $(\mathbf{x}_j, \mathbf{x}_{j+1})$ is minimized. A restricted search volume is used to protect against extreme node perturbations possible in the presence of large velocity gradients. Appropriate radii for each search volume are $\frac{|\mathbf{x}_i - \mathbf{x}_{i-1}|}{2}$ and $\frac{|\mathbf{x}_j - \mathbf{x}_{j+1}|}{2}$, for \mathbf{x}_i and \mathbf{x}_j , respectively. Nonlinear conjugate-gradient is used to perform the minimization, and convergence within the limited search volumes tends to be very quick. Usually no more than three or four line searches are required, each using only a few tens of function evaluations.

Topography is also problematic for ray-bending, though it is handled fairly easily given a good initial path. The simplest approach is to project any perturbation point located above the surface straight downward to the surface, or slightly below it. Though this may not be the minimum time path in regions of rugged topography, it is a sufficient approximation for initial paths with small node spacing. If this solution is found to be inadequate, an optimal search using conjugate-gradient and an appropriate subsurface search volume can be used.

RESULTS

All examples and performance figures shown are based on an ANSI C++ implementation running on a dual processor, 200 MHz Pentium Pro PC. The implementation is used primarily on UNIX platforms but the PC was selected for performance evaluation to minimize variations due to external system load. Graph-theoretic raytracing tests are conducted for all combinations of eight distinct grid-node spacings and forward-star levels of three to seven, except where noted.

Performance with a Constant Gradient Model

A linear gradient model, with velocities given by $V = 3.0 + 0.2 z$, where z = depth, is used to illustrate errors and compare efficiencies of the basic graph-theory traveltimes field algorithm. The linear gradient model allows analytic calculation of traveltimes for comparison (Telford et al., 1976), and provides more realistic assessments of intrinsic geometry related error than a constant velocity model. The traveltimes field volume is selected to be of a typical exploration scale, with dimensions of (10 km x 10 km x 5 km). Traveltimes are computed relative to an assumed source (or receiver) located in the center of the horizontal area, at the surface. Errors are evaluated at one thousand randomly selected locations throughout the model, at depths of greater than one kilometer. A uniform distribution was used to generate all test location coordinates.

Mean absolute error, standard deviation of absolute error, and maximum absolute error for the randomly distributed points are shown in Figure 26. As expected, mean

errors decrease with decreasing node spacing and increasing forward-star level. The maximum percentage error is usually a factor of two or more greater than the mean percentage error. The accuracy shown may be acceptable for some applications, such as rough earthquake location or relative traveltimes studies (Saito, 1989; Nolet and Moser, 1991; Moser et al., 1992), however the maximum absolute errors are too large for detailed tomographic studies or migration operator design.

Before deciding to accept or reject a set of parameters, the issue of mean error, variation in error, and maximum error also needs to be considered within the context of the ultimate application. For tomographic applications, the mean error is related to systematic error that may not significantly alter relative results. The standard deviation of traveltimes errors indicates the average amount of model noise that must be tolerated during inversion. The magnitude of the maximum error relative to the standard deviation of error is related to the potential magnitude of false anomalies that might be generated solely as an artifact of modeling noise.

Prior to selecting a set of parameters for use in traveltimes field computation for a given velocity model, a suite of tests designed to ascertain the relationship between traveltimes error and the primary parameters. These results should be considered along with intrinsic data accuracy and desired resolution when selecting parameters.

The computing resources required to achieve a desired accuracy level represent the second consideration when selecting parameters. The time required to compute a graph-theoretic traveltimes field is at best a linear function of the total number of nodes in the network, assuming no queueing overhead. The number of nodes in the network is

proportional to $(1/\Delta x)^3$, where Δx is the grid node spacing (Figure 28b). On the other hand, the rate of error reduction is sublinear in Δx . This implies that as the node spacing becomes small, computing time rapidly approaches infinity (Figure 28), with little decrease in error.

Given a node spacing and forward-star level, the actual time required to compute a traveltimes field depends primarily on the efficiency with which ray segment traveltimes are computed, and the time required for queue management. The original Dijkstra algorithm is very inefficient, and graph-theory based raytracing first became practical when the heap queue was introduced (Moser, 1991). The quicksort queue marginally improves performance over the heap queue (Cheng and House, 1996), and the unsorted BUCKET was reported to be more efficient than the heap (Klimes and Kvasnicka, 1994). The LIFO stack based bucket queue, introduced in this paper, reduces queueing overhead to the minimum possible level. If zero is used as an origin time, the only operation above pushing an element onto a stack is a single floating point division and an pointer offset dereference. Relative performance of the heap, unsorted BUCKET, and LIFO-stack BUCKET queues are reported (Table 2), for varying numbers of grid nodes and forward-star levels. The same implementation is used, other than the type of queue instantiated at run-time. The LIFO-stack queue is found to be more efficient than the heap, but the difference is only material for large grids. The unsorted BUCKET queue is very inefficient relative to either the heap or the LIFO-stack BUCKET queues. We attribute the previously reported advantage of the unsorted BUCKET to differences in implementation other than queueing.

The above results indicate that approximate traveltimes computation using graph-theory is very efficient, but semi-infinite computing resources are required to achieve high levels of accuracy. Low noise tolerance applications require refinement of raypaths recovered from approximate fields. The nearly irreducible error inherent with graph-theoretic raytracing is primarily due to geometry error arising as a consequence of the finite number of travel paths available. Simply smoothing the approximate raypaths can reduce errors considerably, particularly for coarse grids (Figure 27a and Figure 27b). Simple smoothing is very fast, but does not ensure that the raypath bends in accordance with the principles of ray-geometry physics. Error reduction is limited, and a bending method must be used to achieve arbitrarily small error levels.

High accuracy levels are achieved by perturbing graph-theoretic raypaths using modified pseudo-bending (Figure 27c). It is important to note that the best accuracy is achieved when the most sparsely spaced approximate paths are used. Though dense grids produce more accurate initial raypaths, perturbation of densely sampled approximate raypaths proceeds at a very slow pace. Numerical convergence is detected before the raypath is actually perturbed into a true stationary time geometry. The efficiency of the accuracy of perturbing a sparse initial path relative to that observed for a densely sampled initial path implies that the most sparse grid that does not alias velocity structure should be used to achieve both higher accuracy and reduced computing costs. A study using 100,000 randomly selected end-points for a variety of forward-star levels and grid-node spacings indicates that it is possible to estimate raypaths at a rate of over 1000 rays/second with errors of less than 0.1 millisecond (Figure 29).

Performance with an Exponential Velocity Model

Constant-gradient models are a best case scenario for pseudo-bending. Because of this, we also present an example using a model with exponentially increasing velocity. This model also allows analytic calculation of raypaths, but is characterized by a rapidly increasing velocity gradient. We employ the same parameters and rays used by Grechka and McMechan (1996) (Figure 24). Using a grid-node spacing of 0.15 meters, raypaths are again estimated at over 1000 rays/second with errors of less than 0.1 millisecond. Errors for each raypath are illustrated (Figure 25). These results compare favorably with results reported using a Chebyshev polynomial based raypath representation (Grechka and McMechan, 1996) where raypath estimation with similar error levels was reported to proceed at the rate of a few rays/second using a SPARC 2. It should also be noted that the Chebyshev-polynomial based method requires smooth velocity models.

Performance with a Heterogeneous Grid Model

Grid-based models are often used to describe heterogeneous earth-structure. Fischer and Lees (1993) used a simple, two-anomaly model allowing visual evaluation of raytracer performance. We use a similar procedure. Refined raypath geometries are shown over the velocity model (Figure 30). Shooting in the neighborhood of the refined raypaths failed to locate faster paths, and indicating that the refined paths are within numerical tolerance limits of the minimum time path. Raytracing tests between random endpoints in the model indicate high accuracy at rates approaching 1000 rays/second.

We conclude that ray relaxation rates are not significantly affected by model heterogeneity. The raytracing methods shown here are also used with more complex models for tomographic velocity estimation (Chapter 6, this thesis).

CONCLUSIONS

Accurate, high-performance, raytracing and traveltimes field calculation translates directly into greater ability to test hypotheses within the context of a variety of geophysical tasks. These tasks include testing of raypath/structure modeling, tomographic estimation of velocity structure, optimal hypocenter location/error estimation, and migration operator design. A general use traveltimes field calculator and raypath estimation procedure should be robust, efficient, and accurate. The implementation should allow 'fair' comparison of both existing procedures and new methods as they are proposed. An abstract application framework is capable of providing such a platform for comparison.

Traveltimes calculation and raytracing based on graph-theoretic methods is robust and produces valid approximate results in the presence of arbitrarily complex velocity models. The LIFO-stack BUCKET queue is the most efficient method of managing the queue of fringe-nodes using a graph-theoretic approach. The maximum effective accuracy of graph-theoretic based methods is practically limited, given finite computational resources. Graph-theoretic raypaths and traveltimes may not be useful for applications requiring high accuracy levels.

Raypath relaxation is an efficient method for refining an approximate raypath to a required accuracy level. Pseudo-bending is a special case of raypath relaxation that is limited to use within continuous models with no constant velocity regions. Simple extensions are possible that allow a modified form of pseudo-bending to perform adequately, and efficiently, in the presence of strong velocity field curvature, discontinuities, and regions of constant velocity. This performance is achieved without reference to a specific velocity model representation.

Our general raytracing methodology for arbitrarily complex models is based on refinement of graph-theoretic initial raypaths to required accuracy. Graph-theoretic initial raypaths are efficiently refined to high levels of accuracy using modified pseudo-bending. Correct minimum time raypaths can be recovered after minor velocity model perturbation without traveltimes field recomputation.

ACKNOWLEDGEMENTS

Projects were funded in part by NSF grants EAR91-18090, EAR-94-05471, EAR-95-06379, the Alaska Volcano Observatory, and United States Geological Survey grant 14-08-0001-A0574 administered through the Alaska Volcano Observatory. Scholarships were provided by the Alaska Geophysical Society and Exxon through the SEG Scholarship Foundation .

Computational resources used for this investigation were provided by the University of Alaska Fairbanks Geophysical Institute (Seismology Laboratory), the Alaska Volcano Observatory, the U.S. Geological Survey (Anchorage, AK), and Applied Digital

Technologies. Richard Simon provided key feedback regarding abstract design objectives.

REFERENCES

- Aki, K., and Richards, P., 1980, Quantitative seismology: Theory and methods: W. H. Freeman & Co.
- Bezier, P., 1972, Numerical control--Mathematics and applications: John Wiley & Sons.
- Buzzi-Ferraris, G., 1993, Scientific C++: Addison-Wesley Publishing Company, Inc.
- Cao, S., and Greenhalgh, S., 1993, Calculation of the seismic first-break time field and its ray path distribution using a minimum travelttime tree algorithm, *Geophys. J. Int.*, **114**, 593-600.
- Cerveny, V., 1985, The application of ray tracing to the numerical modelling of seismic wave fields in complex structures, in *Handbook of geophysical exploration*, Section 1, Seismic exploration, Vol. 15A: Geophysical Press, 1-119.
- Cerveny, V., 1987, Ray tracing algorithms in three dimensional laterally varying layered structures, in Nolet, G., Ed., *Tomography in seismology and exploration seismics*: D. Reidel Publ. Co., 99-133.
- Cerveny, V., and Psencik, I., 1983, Gaussian beams and paraxial ray approximation in three-dimensional elastic inhomogeneous media, *J. Geophys.*, **53**, 1-15.
- Chapman, C. H., 1985, Ray theory and its extensions: WKBJ and Maslov seismograms: *J. Geophys.*, **58**, 27-43.
- Cheng, N., and House, L., 1996, Minimum travelttime calculation in 3-D graph-theory: *Geophysics*, **61**, 1895-1898.
- Coultrip, R. L., 1993, High-accuracy wavefront tracing travelttime calculation: *Geophysics*, **58**, 284-292.

- Dijkstra, E. W., 1959, A note on two problems in connection with graphs: *Numer. Math.*, **1**, 269-271.
- Farra, V., Virieux, J., and Madariaga, R., 1989, Ray perturbation theory for interfaces, *Geophys. J. Internat.*, **99**, 377-390.
- Fischer, R., and Lees, J. M., 1992, 3-D raytracing for tomographic inversion and earthquake relocation: A comparison of methods: *Eos Trans. AGU*, **73**, 198.
- Fischer, R., and Lees, J. M., 1993, Shortest path ray tracing with sparse graphs: *Geophysics*, **58**, 987-996.
- Fletcher, R., 1987, *Practical methods of optimization*: John Wiley and Sons.
- Gjoystdal, H., Reinhardsen, J. E., and Ursin, B., 1984, Traveltime and wavefront curvature calculations in three-dimensional inhomogeneous layered media with curved interfaces: *Geophysics*, **49**, 1466-1494.
- Grechka, V. Y., and McMechan, G. A., 1996, 3-D two-point ray tracing for heterogeneous, weakly transversely isotropic media: *Geophysics*, **61**, 1883-1894.
- Hanyga, A., 1988, Numerical methods for tracing rays and wavefronts, in Doornbos, D. J., Ed., *Seismological algorithms: Computational methods and computer programs*: Academic Press.
- Hole, J. A., 1992, Nonlinear high-resolution three-dimensional seismic traveltime tomography, *J. Geophys. Res.*, **97**, 6553-6562.
- Julian, B. R., and Gubbins, D., 1977, Three dimensional seismic ray tracing: *J. Geophys.*, **43**, 95-114.
- Keller, H. B., and Perozzi, D. S., 1983, Fast seismic ray tracing: *SIAM J. Appl. Math.*, **43**, 981-992.
- Klimes, L., and Kvasnicka, M., 1994, 3-D network ray tracing: *Geophys. J. Int.*, **116**, 726-738.

- Knuth, D. E., 1968, *Fundamental Algorithms*, vol. 1 of: *The art of computer programming*, Addison-Wesley.
- Langan, R. T., Kerche, I., and Cutler, R. T. 1985, Tracing of rays through heterogeneous media: An accurate and efficient procedure: *Geophysics*, **50**, 1456-1465.
- Lees, J. M., and Shalev, E., 1992, On the stability of P-wave tomography at Loma Prieta: A comparison of parameterizations, linear and nonlinear inversions: *Bull. Seis. Soc. Am.*, **82**, 1821-1839.
- Liu, Q., 1991, Solution of the 3-D eikonal equation by an expanding wavefront method: 59th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1488-1491.
- Lyche, T., and Morken, K., 1986, Making the Oslo algorithm more efficient, *SIAM*, **23**, 663-675.
- Moser, T. J., 1991, Shortest path calculation of seismic rays: *Geophysics*, **56**, 59-67.
- Moser, T. J. 1992, The shortest path method for seismic ray tracing in complicated media, Ph.D. thesis, Institute of Geophysics, Utrecht University, Netherlands.
- Moser, T. J., 1993, Calculation of the seismic first-break time field and its ray path distribution using a minimum travelttime tree algorithm, *Geophys. J. Int.*, **114**, 593-600.
- Moser, T. J., van Eck, T., and Nolet, G., 1992, Hypocenter determination in strongly heterogeneous earth models using the shortest path method: *J. Geophys. Res.*, **97**, 6563-6572.
- Muller, G., 1984, Efficient calculation of Gaussian-beam seismograms for two-dimensional inhomogeneous media, *Geophys. J. Roy. Astr. Soc.*, **79**, 153-166.
- Nakanishi, I., and Yamaguchi, K. 1986, A numerical experiment on nonlinear image reconstruction from first-arrival times for two-dimensional island arc structure: *J. Phys. Earth*, **34**, 195-201.

- Nelder J. A., and Mead, R., 1965, A simplex method for function minimization: *Computer J.*, **7**, 308-313.
- Nolet, G., and Moser, T. J., 1991, Traveltimes in a 3-D earth and a new look at the *S* discrepancy: *EOS Trans. AGU*, **72**, 316.
- Pereyra, V., 1988, Two-point ray tracing in complex 3-D media: 58th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1056-1060.
- Pereyra, V., Lee, W. H. K., and Keller, H. B., 1980, Solving two-point seismic ray tracing problem in a heterogeneous medium, Part I: A general adaptive finite difference method: *Bull. Seis. Soc. Am.*, **70**, 79-99.
- Podvin, P., and Lecomte, I., 1991, Finite-difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools: *Geophys. J. Internat.*, **105**, 271-284.
- Polak, E., 1971, *Computational methods in optimization*: Academic Press.
- Prothero, W. A., Eickemeyer, J., and Taylor, W. J., 1988, A fast two-point raytracing algorithm using a simple step search method: *Bull. Seism. Soc. Am.*, **78**, 1190-1198.
- Pusey, L., and Vidale, J., 1991, Accurate finite-difference calculation of WKBJ traveltimes and amplitudes: 61st Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1513-1516.
- Qin, F., Luo, Y., Olsen, K., Cai, W., and Schuster, G., 1992, Finite-difference solution of the eikonal equation: *Geophysics*, **57**, 478-487.
- Qin, F., and Schuster, G., 1993, First-arrival traveltime calculation for anisotropic media: *Geophysics*, **58**, 1349-1358.

- Saito, H., 1989, Traveltimes and raypaths of first arrival seismic waves: Computation method based on Huygens's principle: 59th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 244-247.
- Sambridge, M. S., and Kennett, B. L. N., 1990, Boundary value ray tracing in heterogeneous medium: A simple and versatile algorithm: Geophys. J. Internat., **101**, 157-168.
- Schneider, W. A., Ranzinger, K. A., Balch, A. H., and Kruse, C., 1992, A dynamic programming approach to first arrival traveltime computation in media with arbitrarily distributed velocities, Geophysics, **57**, 39-50.
- Stone, B. D., and Forbes, G. W., 1990, Optimal interpolants for Runge-Kutta ray tracing in inhomogeneous media, J. Opt. Soc. Am., **7**, 248-254.
- Stroustrup, B., 1991, The C++ programming language: AT&T Bell Telephone Laboratories, Inc.
- Stroustrup, B., and Ellis, M. A., 1994, The annotated C++ reference manual, AT&T Bell Laboratories, Inc.
- Telford, W. M., Geldart, L. P., Sheriff, R. E., and Keys, D. A., 1976, Applied geophysics: Cambridge University Press.
- Um, J., and Thurber, C., 1987, A fast algorithm for two-point seismic ray tracing: Bull. Seis. Soc. Am., **77**, 972-986.
- Van Trier, J., and Symes, W., 1991, Upwind finite-difference calculation of traveltimes: Geophysics, **56**, 812-821.
- Vidale, J., 1988, Finite-difference calculation of traveltimes: Bull. Seis. Soc. Am., **78**, 2062-2076.
- Vidale, J., 1990, Finite-difference calculation of traveltimes in three dimensions: Geophysics, **55**, 521-526.

- Virieux, J., 1991, Fast and accurate ray tracing by Hamiltonian perturbation: J. Geophys. Res., **96**, 579-594.
- Virieux, J., Farra, V., and Madariaga, R., 1988, Ray tracing in laterally heterogeneous media for earthquake location: J. Geophys. Res., **93**, 6585-6599.
- Wattus, N., 1991, 3-D finite-difference traveltimes in complex media: 59th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1106-1109.
- Zhao, D., 1990, A tomographic study of seismic structure in the Japan Islands: Ph. D. thesis, Tohoku University, Japan.

**Comparison of Pseudo-Bending Performance using
Linear Slowness and Linear Velocity
Segment Traveltime Estimation
for Differing Convergence Thresholds and Exaggeration Factors**

p	F	<i>Error</i> (ms) reported by Um & Thurber	<i>Error</i> (ms) Linear Slowness	<i>Error</i> (ms) Linear Velocity	<i>Cost</i> (ms) Linear Slowness	<i>Cost</i> (ms) Linear Velocity
0.1	1	144.2	40.47	33.24	0.9	0.5
0.01	1	67	17.69	9.11	3.8	2.2
0.001	1	18.6	6.11	2.34	8.1	5
0.001	1.3	12.1	3.29	1.42	10	5.9
0.001	1.6	6.8	2.65	1.37	19.7	12.8
0.001	1.9	2.5	9.98	7.48	28	22.4
0.0001	1	6.2	1.8	0.77	33	19.5

Table 1: Comparison of pseudo-bending performance using linear slowness versus linear velocity segment traveltime estimation.

The velocity model used is $V = 4.0 + 0.2 z$. The assumed source is located at a horizontal distance of one hundred kilometers and a depth of ten kilometers from the receiver. The analytic traveltime is 14.664 seconds. Three different errors are reported for each combination of convergence threshold, p , and exaggeration factor, F . Column three is the original error reported by Um and Thurber (1987) using the original pseudo-bending algorithm. The error reported in column four is computed using the end to end relaxation schedule, and considerable improvement is observed. Significant additional error reduction is possible by using the linear velocity variation equation (29), as is shown in column five. Comparison of the last two columns shows the linear velocity variation equation (29) always results in more rapid convergence.

<i>Grid-node Spacing (km)</i>	<i>Forward-star level</i>	<i>Total Number Of Nodes</i>	<i>Bucket Queue Comp. Time (s)</i>	<i>Heap Queue Comp. Time (s)</i>	<i>Dynamic Binned Queue Comp. Time (s)</i>	<i>Max. Error %</i>
0.500	3	4851	1.031	0.735	0.719	3.4178
0.417	3	8125	1.985	1.297	1.25	3.2195
0.357	3	12615	3.782	2.031	1.984	3.1641
0.294	3	22050	8.625	3.703	3.516	2.5382
0.238	3	40678	24.922	7.047	6.563	2.5663
0.185	3	84700	94.594	15.25	13.875	2.438
0.135	3	213750	502.359	40.453	35.781	2.4051
0.085	3	849660	N/C	169	145.266	2.3622
0.500	4	4851	1.703	1.422	1.375	3.4178
0.417	4	8125	3.265	2.453	2.375	1.9869
0.357	4	12615	6.031	3.906	3.781	1.7852
0.294	4	22050	13.156	7.031	6.703	2.1472
0.238	4	40678	36.063	13.453	12.61	1.6094
0.185	4	84700	134.453	29.234	26.734	1.4843
0.135	4	213750	672.109	78.937	69.407	1.4642
0.085	4	849660	N/C	336.11	280.875	1.4495
0.500	5	4851	2.672	2.5	2.282	3.4178
0.417	5	8125	4.984	4.219	3.985	1.951
0.357	5	12615	9.094	6.875	6.296	1.3019
0.294	5	22050	19.375	12.75	11.234	1.322
0.238	5	40678	53.672	23.718	21.25	1.5087
0.185	5	84700	183.375	51.829	45.468	0.948
0.135	5	213750	877.063	137.687	118.594	0.889
0.085	5	849660	N/C	547.516	485.687	0.87
0.500	6	4851	4.047	3.766	3.641	3.4178
0.417	6	8125	7.547	6.562	6.297	1.951
0.357	6	12615	13.547	10.687	10.11	1.2292
0.294	6	22050	28.844	19.344	18.14	1.0932
0.238	6	40678	75.016	37.438	34.578	0.771
0.185	6	84700	245.25	81.36	74.14	0.939
0.135	6	213750	1110.656	215.016	194.796	0.619
0.085	6	849660	N/C	894.843	796.406	0.608
0.500	7	4851	5.407	5.125	4.89	3.4178
0.417	7	8125	10	8.891	8.563	1.951
0.357	7	12615	17.922	14.516	13.828	1.2292
0.294	7	22050	38.047	26.562	24.844	1.0932
0.238	7	40678	96.281	51.453	47.547	0.7
0.185	7	84700	301.813	111.875	102.032	0.701
0.135	7	213750	1305.297	296.89	267.704	0.593
0.085	7	849660	N/C	1237.703	1100.671	0.583

Table 2: Comparison of traveltime field calculation costs and associated maximum errors for various forward-star levels and grid node spacings.

Computation times reported in the table span almost four orders of magnitude while the maximum errors span less than one order of magnitude. Node-spacing and forward-star parameters must be carefully selected within the context of required accuracy to avoid large computation times that might be avoided.

Tests were computed using a 200 MHz Pentium PC. A constant gradient velocity model with $V = 3.0 + 0.2 z$ (km/s) was used over a (10 km x 10 km x 5 km) volume.

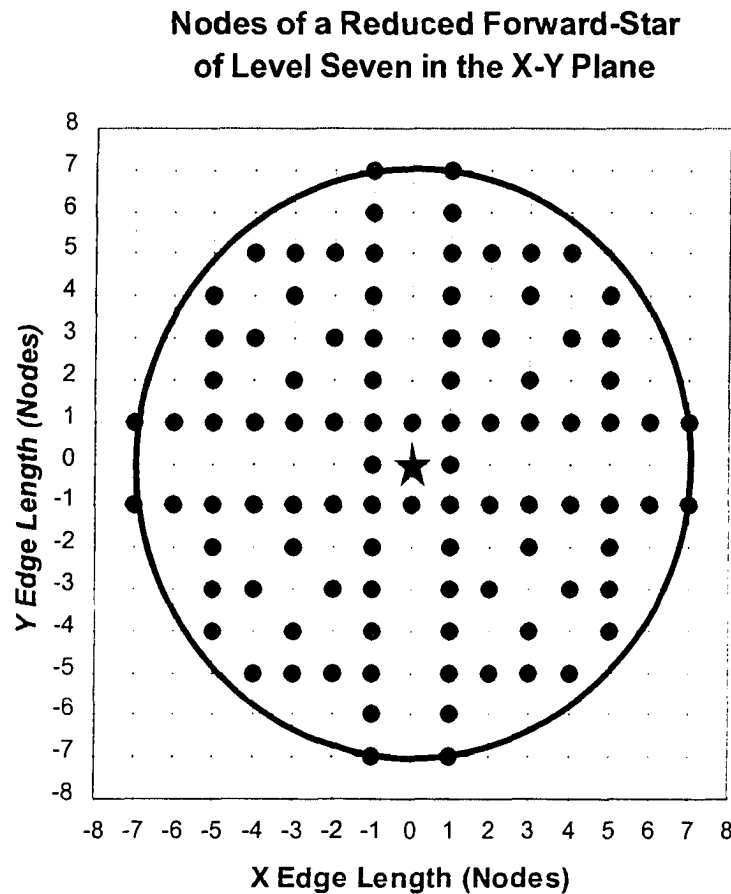


Figure 20: Nodes corresponding to edges of a reduced, 3-D, forward-star, in the x-y plane.

Many edges have been removed without altering the maximum angular gap between edges. The edges of cubical forward-stars include all grid-line intersections with node indices less than or equal to seven. Spherical forward-stars only include the grid-line intersections within the large circle. Optimally reduced spherical forward-stars have all long edges removed that do not increase the maximum angular gap between edges. A level seven spherical forward-star based on a regular 3-D grid has a maximum angular gap between edges of a degrees.

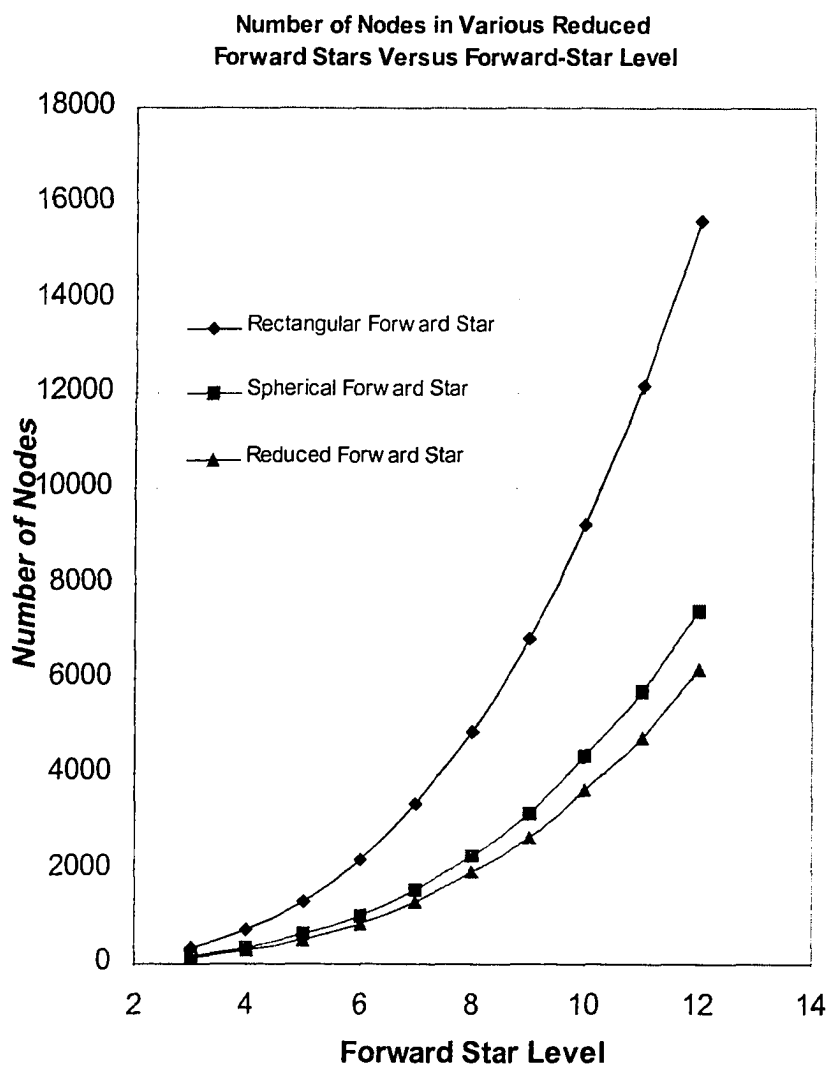


Figure 21: Comparison of edge counts for reduced and unreduced forward-stars. Spherical forward-stars contain fewer edges than rectangular forward-stars of the same level. Further reduction is possible by eliminating edges that do not decrease the maximum angular gap between edges.

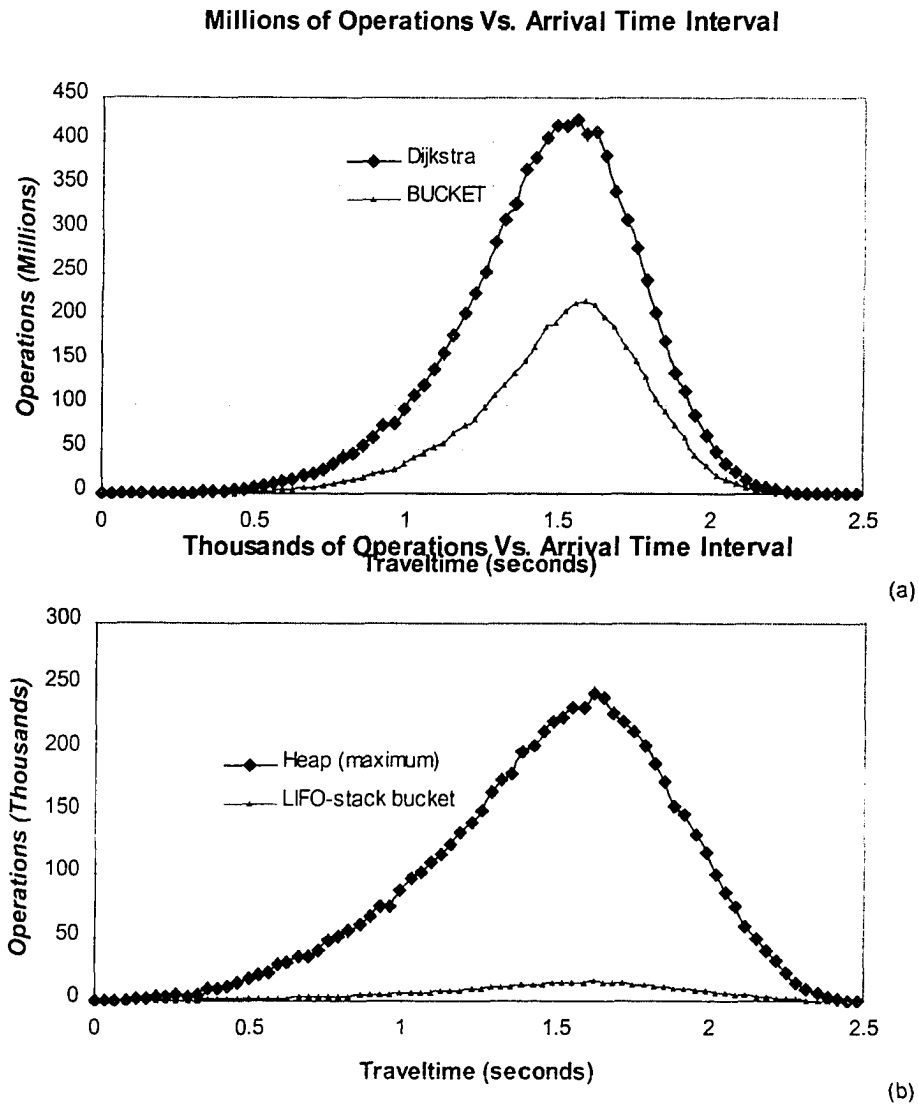


Figure 22: Number of required queueing operations versus traveltime interval for various proposed queues.

(a) The number of operations empirically measured for Dijkstra and BUCKET queues. (b) The maximum number of operations possible for a heap queue and the empirically measured number of operations required for a LIFO-stack bucket queue.

Note that there are **three orders of magnitude difference** between the vertical scales. Results are computed for a level seven forward-star and, 213,750 grid nodes. The grid interval is 0.135 km, and the target volume is (10 km x 10 km x 5 km). The source was placed at (0,0,0). The linear gradient velocity model is used with $V = 3.0 + 0.2 \cdot \text{depth}$, over a volume of (10 km x 10 km x 5 km).

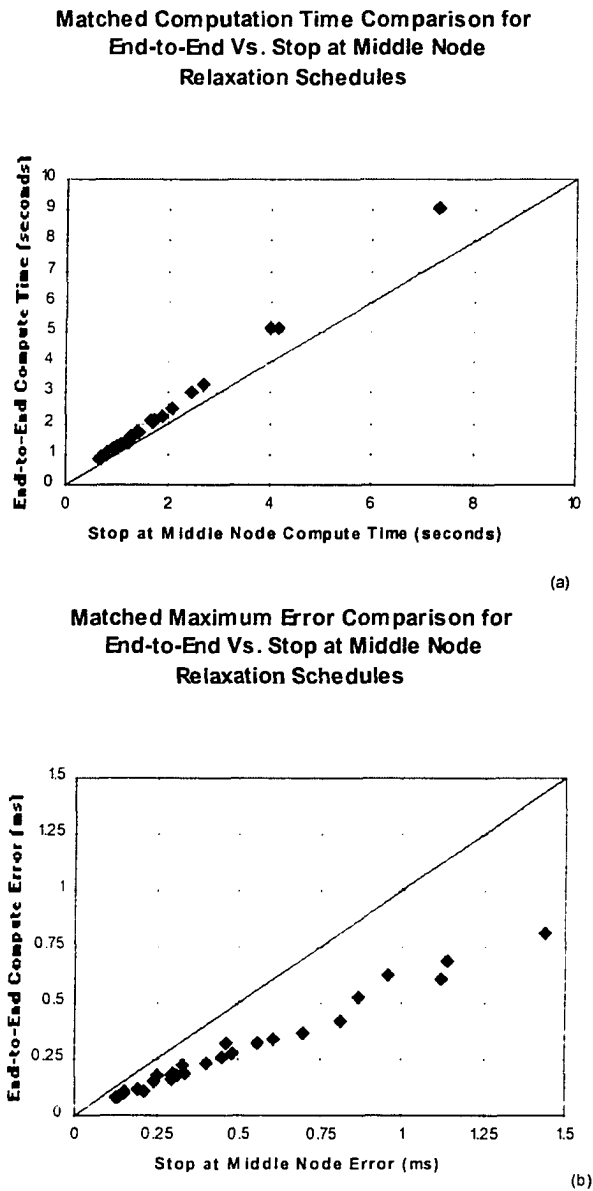


Figure 23: Maximum error and computation cost for various relaxation schedules using one-thousand rays and a constant gradient velocity model.

Maximum error and total execution times are shown for matched initial raypaths. The initial raypaths are derived from different approximate traveltimes fields, but with the same randomly selected endpoints. The traveltimes fields covered a (10 km x 10 km x 5 km) volume. A constant gradient velocity model, $V = 3.0 + 0.2z$, was used.

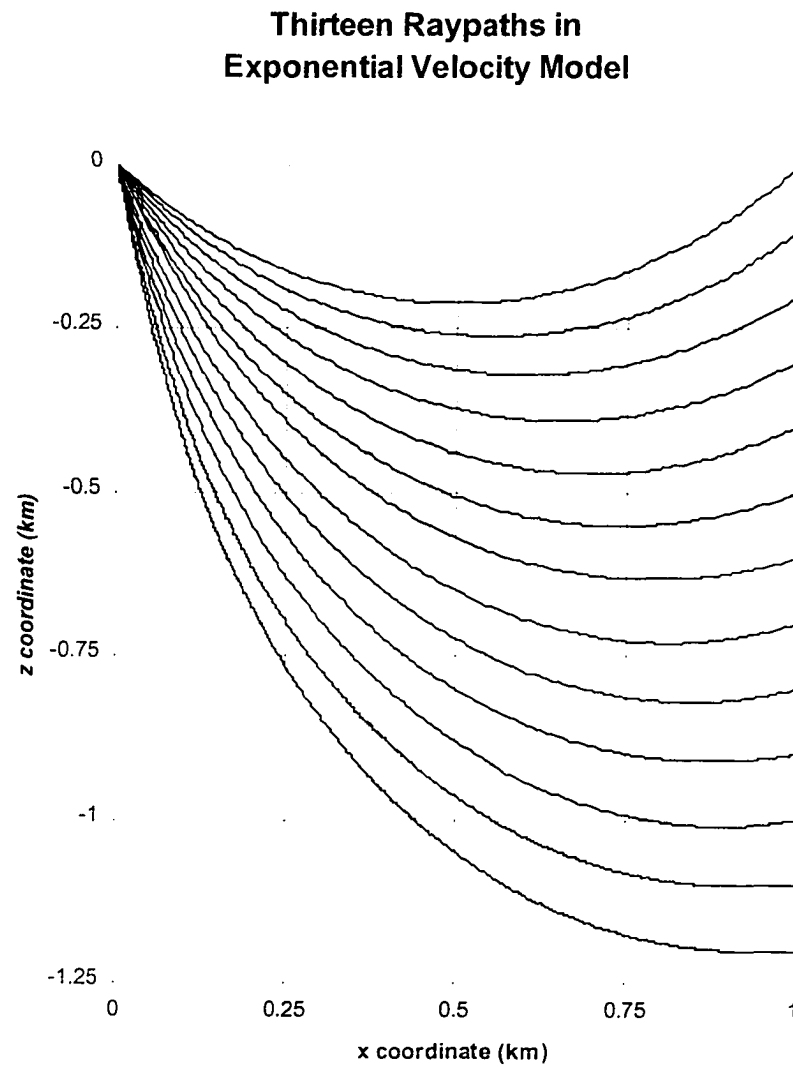


Figure 24: Thirteen rays in an exponential velocity model.

Raypaths used for testing the algorithm in a velocity model with $V = 1.5 \exp(-1.5z)$, where $z = -\text{depth}$. Grechka and McMechan (1996) also use the same model and raypaths.

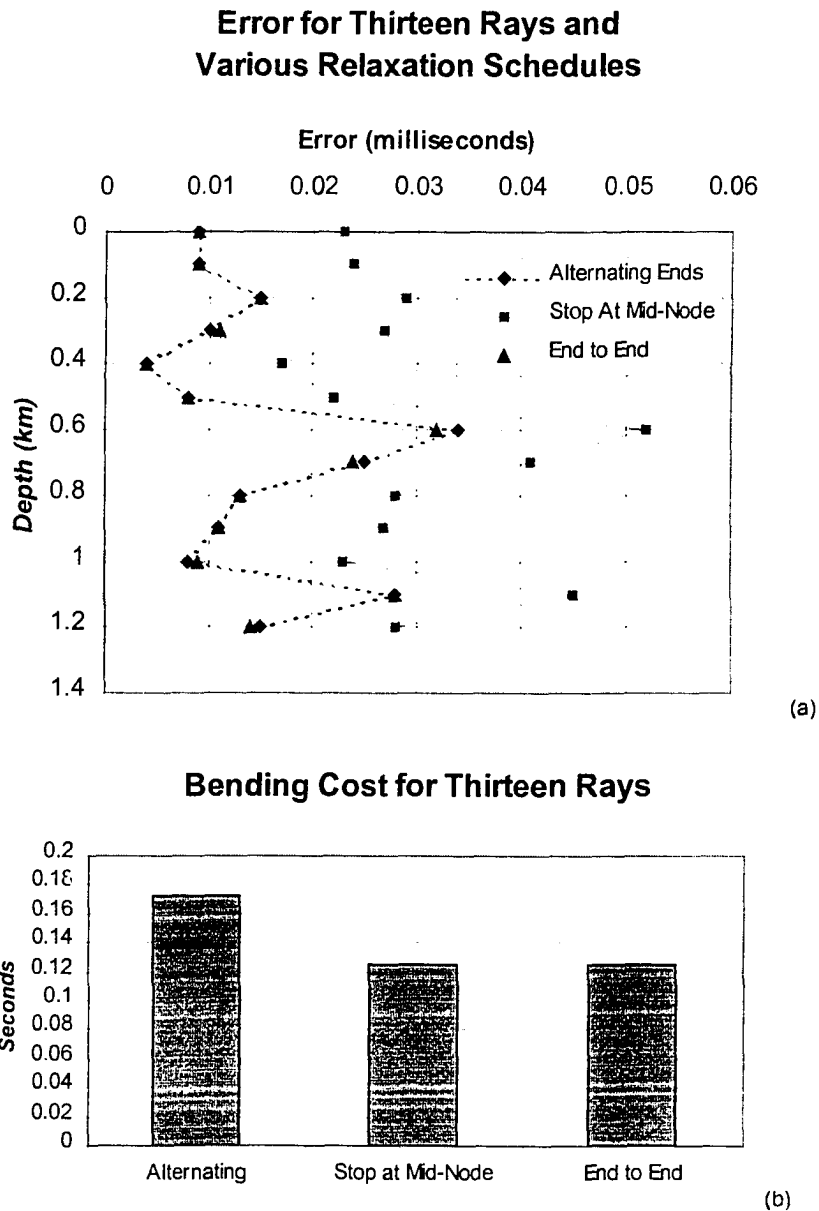


Figure 25: Error and computation cost for various relaxation schedules using thirteen rays and an exponentially increasing velocity model.

The end to end relaxation schedule produces smaller errors with no additional computing cost relative to the stop at mid-node schedule. The alternating end schedule also produces small errors, but with larger costs.

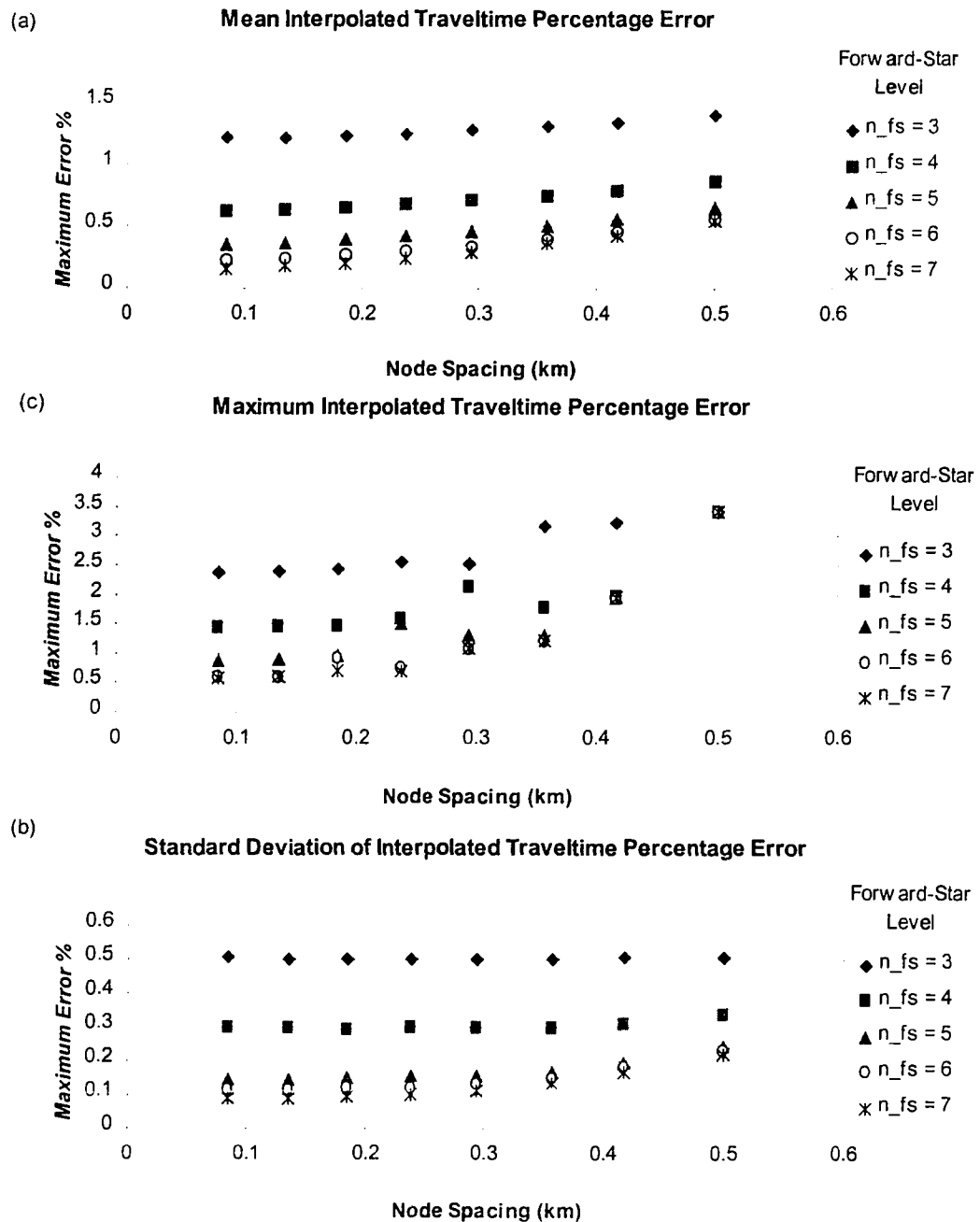


Figure 26: Traveltime percentage errors for 1000 random locations in a constant gradient model.

Traveltimes are calculated using graph-theory and tri-linear interpolation. Grid node spacings of (500 m, 417 m, 357 m, 294 m, 238 m, 185 m, 84 m), and forward-star levels (3 - 7) are illustrated. (a) Mean percent error. (b) Standard deviation of percent error. (c) Maximum percent error.

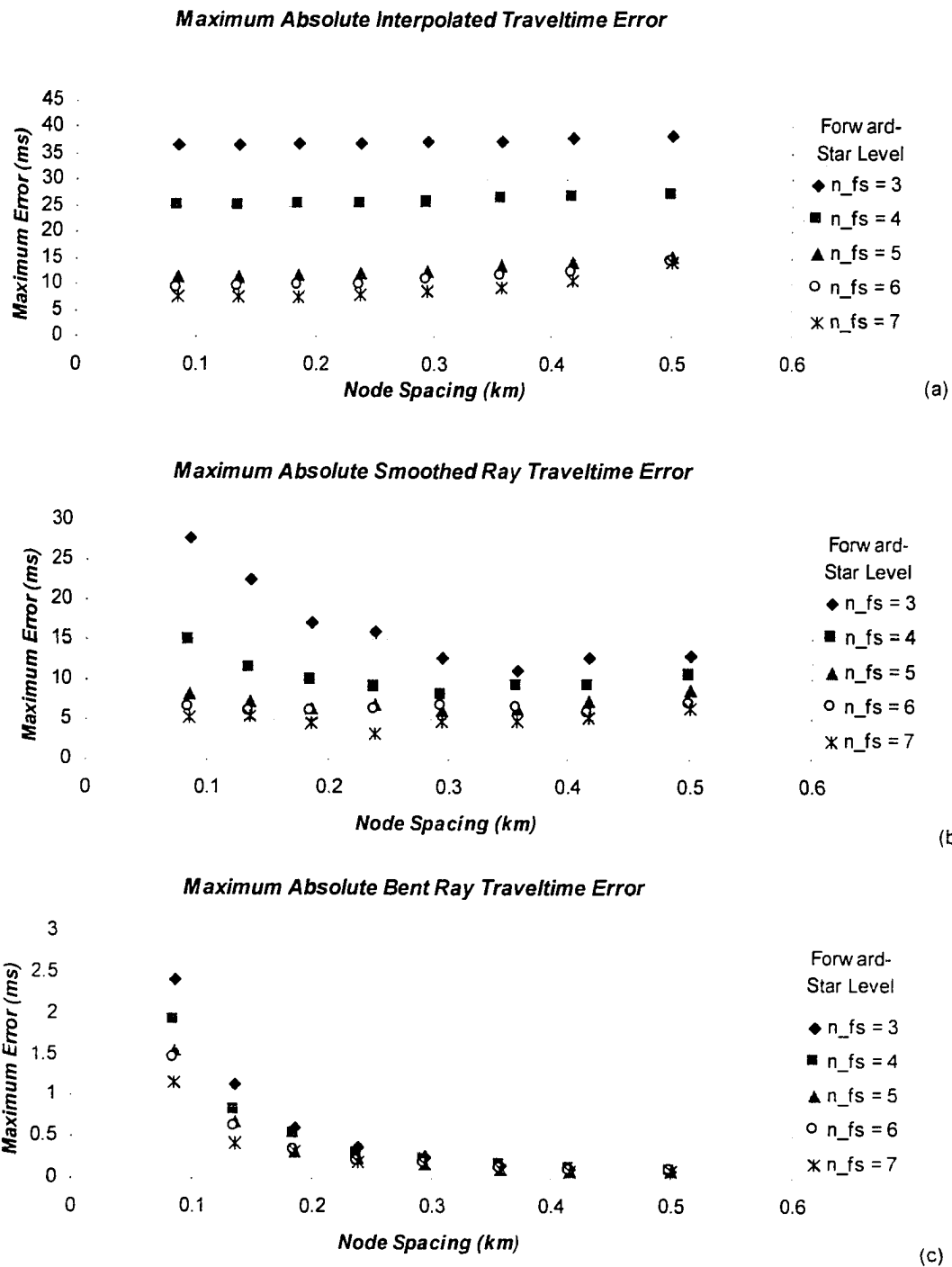


Figure 27: Maximum absolute errors for 1000 random locations in the constant gradient model.

(a) Interpolated only. (b) After ray smoothing. (c) After ray-bending.

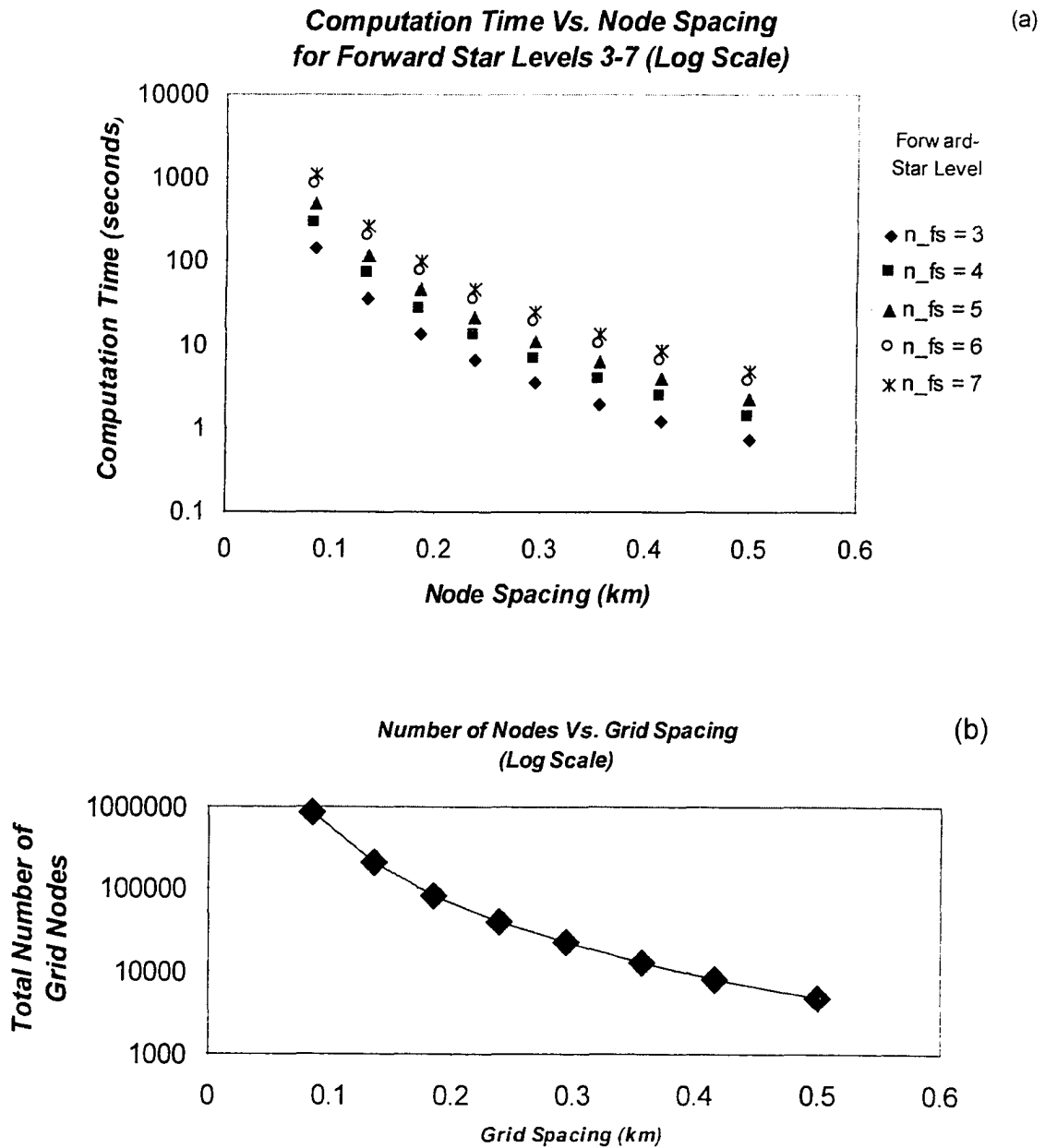


Figure 28: Computing cost and number of nodes as a function of grid spacing of the (10 km x 10 km x 5 km) study volume.

a) Computation time ranges over three orders magnitude for the parameters tested. Errors over the same range only change by a factor of four; b) The total number of grid nodes increases rapidly as the grid spacing decreases, given a constant target volume. Computation time is approximately a linear function of the number of grid nodes for LIFO-stack BUCKET queue implementations.

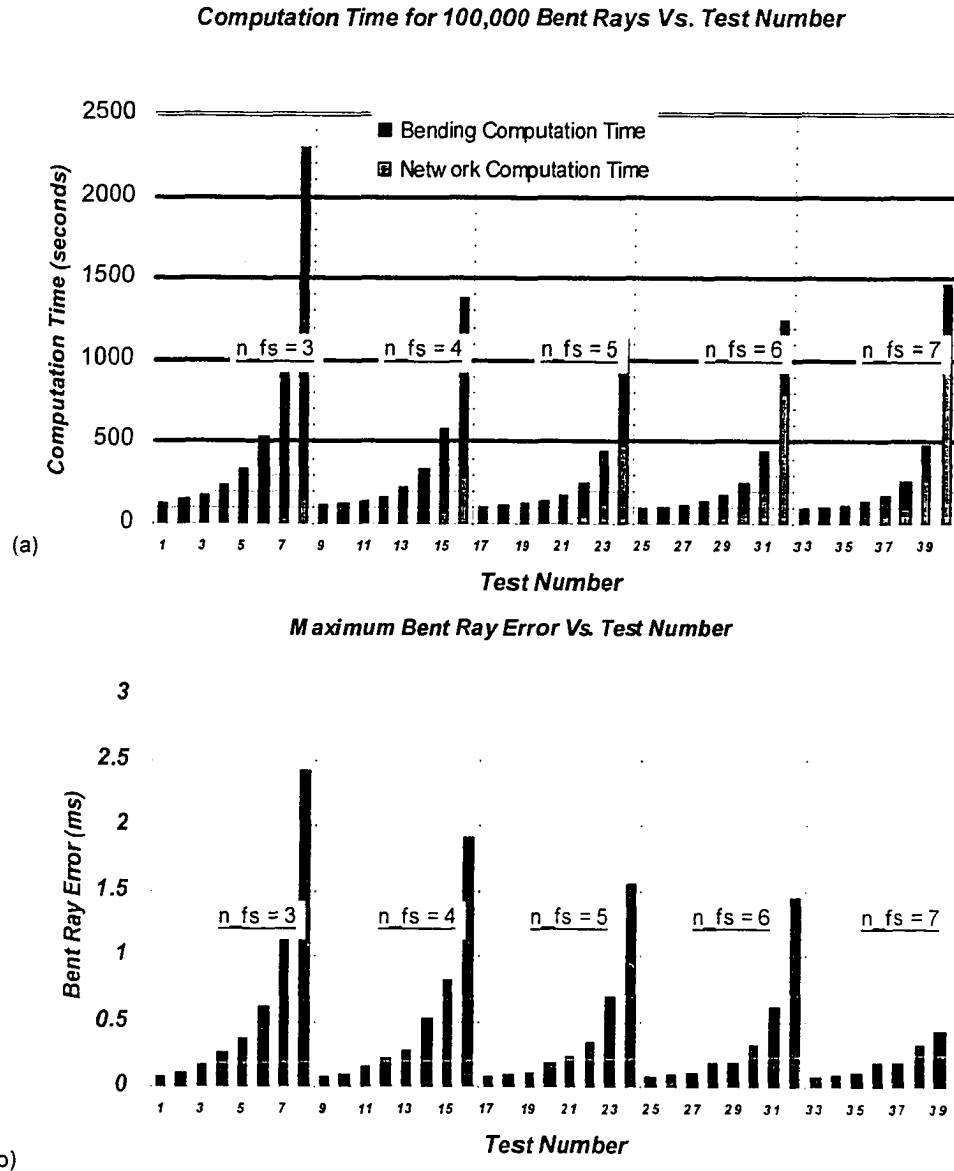


Figure 29: Computational cost and maximum error for 100,000 randomly distributed rays.

(a) Computation time versus trial number. (b) Maximum error versus trial number. Maximum errors of less than 0.1 millisecond and net raypath estimation rates in excess of 1000 rays/second are possible (200 MHz PC). Trials are performed for forward-star levels (3 - 7). Results are shown for each forward-star and grid-node spacings of (500 m, 417 m, 357 m, 294 m, 238 m, 185 m, and 84 m). Smaller trial numbers are associated with larger node spacings within each forward-star grouping. Better accuracy and shorter computation time is associated with sparse grid-node spacing for all forward-star levels.

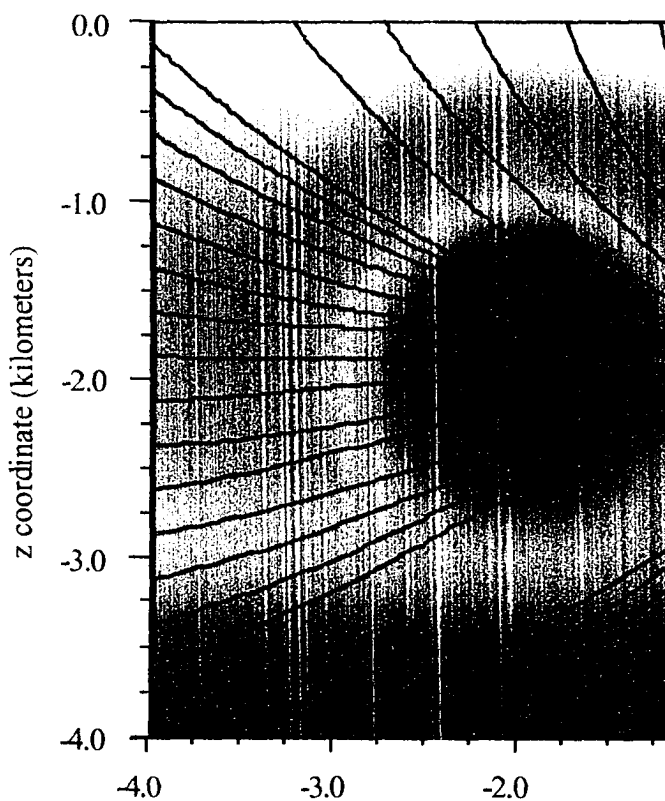
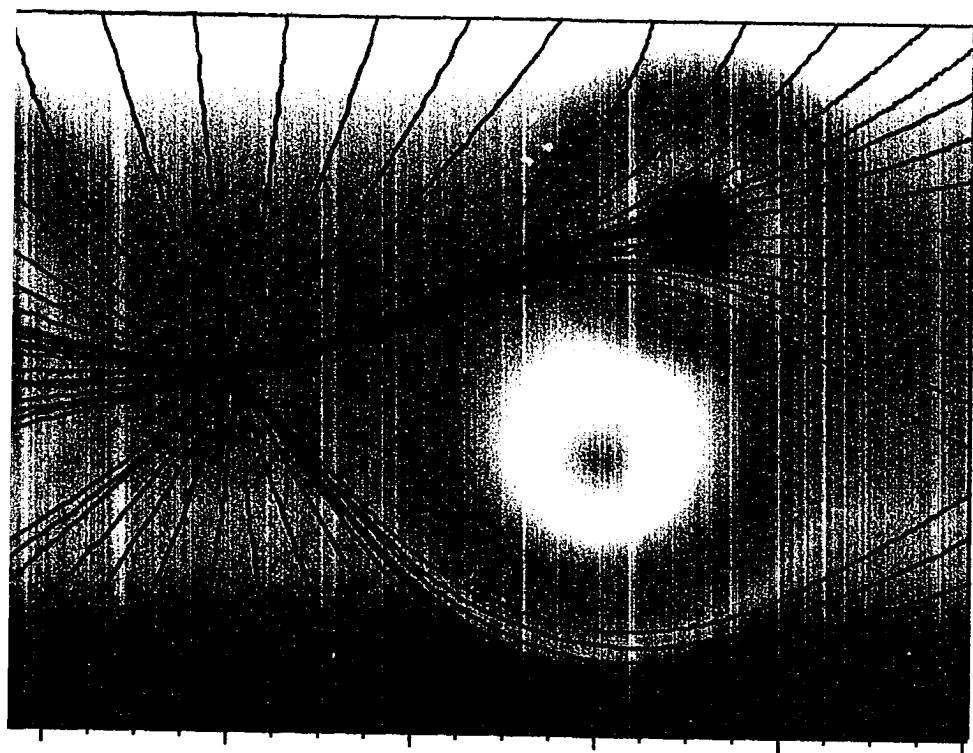
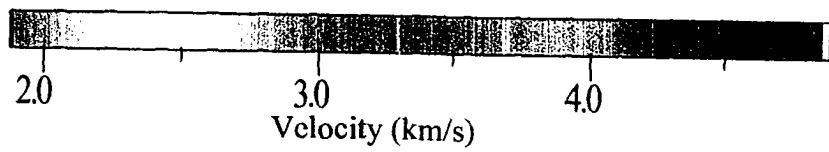


Figure 30:

Raypaths in a heterogeneous, grid-based model. A single source was assumed in the center of the model, and rays are calculated for sixty points around the perimeter. Ray concentration in fast regions and avoidance of slow regions is apparent.



x coordinate (kilometers)



CHAPTER 5

Analysis of the Spring 1996 Earthquake

Swarm at Akutan Volcano, Alaska

ABSTRACT

An energetic earthquake swarm and extended decay sequence occurred during the spring of 1996 at Akutan Volcano, Alaska. The swarm began with two episodes of seismicity, each of which lasted less than 24 hours, separated by a twelve-hour period of relative quiescence. More felt events occurred during these two episodes than in any other swarm in the Global Volcano Earthquake Swarm Database (Benoit and McNutt, 1996).

Both episodes reached rates of more than twenty events per hour within four hours of episode initiation. Four individual earthquakes of the swarm are estimated to be larger than $M_L = 5.0$, and we estimate the total energy released to be greater than a single $M_w = 5.9$ event. Frequency-magnitude law b-values, derived from coda duration magnitudes calibrated to M_L , indicate a Gutenberg-Richter b-value of 1.05 for $M_L < 4.5$. Events larger than $M_L = 4.5$ appear to be associated with a somewhat larger b-value, suggesting the possibility of a greater rupture process dimension for the larger events.

An emergency local array was deployed shortly after the two energetic episodes. Elevated event counts were observed throughout a four month long decay sequence. The time distribution of events during the decay sequence is consistent with the modified Omori Law, suggesting that decay sequence seismicity was associated with passive stress field relaxation. We estimate a very high Omori Law p-value of approximately 2.5, which is consistent with the interpretation of a high-temperature and/or highly fractured hypocentral volume.

Analysis of hypocenter distribution and phase arrival patterns during the decay sequence show that most of the observed seismicity could be associated two to three, closely spaced, nearly point-source, volumes. The hypocenter concentration region is located approximately eight kilometers east of the caldera at a depth of less than four kilometers. A linear feature apparent on maximum likelihood epicenter point plots is similar to the intrinsic epicentral ambiguity function associated with the dominant arrival patterns and the receiving array geometry. We speculate that the initial swarm was associated with two pulses of magmatic intrusion near the hypocenter concentration areas. Hypocenter estimation errors are too large to support a detailed interpretation of intrusion geometry.

INTRODUCTION

An energetic earthquake swarm and extended decay sequence occurred during the spring of 1996 at Akutan Volcano, Alaska. Akutan Volcano is a tholeiitic stratovolcano located on west-central Akutan Island (Figure 31) about 1,280 kilometers southwest of

Anchorage. Akutan is one of the most active volcanoes in the Aleutian arc, with at least 27 eruptions reported since 1790 (Miller et al., 1997). Eruptive activity presents a hazard to the nearby City of Akutan, a large seafood processing plant, and air traffic of over 10,000 people per day, plus cargo, in the vicinity.

Earthquake swarms are generally defined as a sequence of earthquakes with sources that are closely clustered in both time and space, with no outstanding single shock (Mogi, 1963). The 1996 Akutan earthquake swarm consisted of two episodes of energetic seismicity, separated by a forty hour hiatus, and a four month long decay sequence. The two episodes might be considered as separate swarms, but we refer to them as episodes of the same swarm for the purposes of this paper. They were closely spaced in time and space, and were clearly related to the same source process.

Felt ground motion was sufficiently strong that the City of Akutan evacuated a significant percentage of the residents, and many processing plant employees fled the island. The estimated number of felt events at Akutan (3000) is the largest in the Global Volcanic Earthquake Swarm Database (Benoit and McNutt, 1996). The database includes observations from over 700 swarms.

Alaska Volcano Observatory monitoring efforts during the decay sequence focused on monitoring both the temporal and spatial stability of the sequence, with particular emphasis on hazard assessment. The first objective of real-time volcano hazard assessment is to detect indicators of restlessness, active process rejuvenation, and known eruption precursory phenomena. Very few eruptions occur without anomalous precursory seismicity (McNutt, 1996). Potential alert indicators include an increase in

seismicity rate, total seismic energy, volcanic tremor, a change in event-type ratios (VT count / LP count), or a shift in hypocenter concentration volume. At Akutan, one LP event was recorded during the initial episodes, and nine others were observed during the decay sequence. None seemed to be directly associated with the volcano, and were most likely related to non-volcanic processes, such as small submarine landslides. As a consequence, Akutan decay sequence monitoring focused primarily on event counts, magnitudes, and hypocenter location.

The purpose of this paper is to describe the Akutan seismic swarm and decay sequence within the context of seismic monitoring / hazard evaluation objectives. We are specifically interested in techniques for monitoring the temporal and spatial characteristics of decay sequence seismicity at remote and/or marginally monitored locations. Limited-aperture networks, poorly constrained velocity structure, and rugged topographic interference with shallow raypaths are common challenges that must be addressed by data analysis systems used in monitoring seismic crises at remote volcanoes. The procedures employed at Akutan may be useful at other locations.

In the first section we introduce the available data volume and the chronology of the swarm. A coda duration magnitude scale is developed, and energy release rates are presented. We estimate that the total energy released during the two initial episodes was equivalent to that associated with a single $M_L = 5.9$ event. Post-energetic episode event counts are found to be consistent with the modified Omori relationship, suggesting a passive material response after the initial stress perturbation. Relative arrival times are used to establish that the source volume was spatially stable throughout the decay

sequence. A very simple velocity model is developed assuming a linear gradient and constant V_p/V_s ratio. The V_p/V_s estimate is made without hypocentral parameter estimates or traveltimes calculations. The simple velocity model is used in conjunction with L_1 norm, L_2 norm, and Tarantola-Valette L_2 cost functions to compute hypocentral likelihood density functions (HLDFs). HLDF's are used to qualitatively assess the magnitude of intrinsic source location estimation errors. We use this result, together with the raw arrival time information, to conclude that most of the decay sequence seismicity could be associated with two to three, closely spaced, point-sources.

SWARM DATA AND ANALYSIS

Chronology of the Akutan Swarm, Available Data, and Operational Objectives

Anomalous 1996 seismicity at Akutan was first instrumentally recorded on the morning of March 11 at 04:32 GMT at station DTN, located approximately 260 km northeast of Akutan (Figure 31). The detection threshold is approximately $M_L = 3.5$. Within four hours of inception swarm, intensity reached a rate of over forty events per hour of $M_L \geq 3.5$, recorded at DTN (Figure 32). This is a conservative estimate of the total event count because only events with identifiable and consistent $S-P$ times were included. In addition, overlapping event coda obscured some events. Residents of the City of Akutan (located 13 km from the summit) reported "continuous ground motion", punctuated by frequent strong shocks. The total number of felt events is estimated to be greater than 3000 (R. Rozier, written communication).

About seven hours after the beginning of the episode, seismicity declined to a rate of 10-20 events/hour. The decline continued and only 5 total events were recorded on March 13. Then on March 14, after more than 40 hours of relative quiescence, energetic seismicity resumed at a rate of over 20 events/hour ($M_L \geq 3.5$) for a period of approximately 17 hours (Figure 32). Though reports of frequent felt ground motion continued, we regard seismicity after March 14 as part of the four month long decay sequence.

The two initial episodes of energetic seismicity were similar in that both evolved from a quiet state to over 20 events/hour in less than two hours. This implies a rapid source activation process in both cases. The first episode included an early seismicity rate of more than 40 events/hour, while the second episode never exceeded 30 events/hour. Both counts use the $M_L \geq 3.5$ at DTN threshold. The first episode flattened at a rate of 12-20 events/hour for 10 hours prior to subsiding over a 4 hour interval, while the second episode dropped directly from its peak of 29 events/hour to a level of a few events/day in a six hour period. All the events recorded at DTN and attributed to Akutan had identical *S-P* times, within phase identification error estimates.

No local seismic instrumentation was in place when the swarm began. Neither precise location nor focal mechanism estimates are possible for initial swarm earthquakes. The Alaska Volcano Observatory deployed an emergency seismic array during the eight-day period after the last Akutan event was observed at DTN. The array consisted of five vertical, short-period, seismometers and one broad-band, three-component, instrument. Weather and timing constraints did not permit optimal array

configuration. Two vertical seismometers and the three-component instrument were co-located, resulting in four unique station locations (Figure 33), and a roughly linear array geometry. The network became operational on March 23, and was used to monitor the decay sequence.

Early Episode Magnitudes, Seismic Energy, and b-Values

The regional network in place at the time of swarm inception only permitted estimation of forty-seven standard magnitudes during the early episodes. A total of 299 distinct coda lengths were manually measured, however. From these data a homogeneous set of magnitude estimates and corresponding coda duration measurements were identified for thirty-eight events. This data volume permits derivation of a local duration magnitude scale.

The concept of duration magnitude is based on the idea that a linear relationship exists between coda duration and seismic moment. This might be expected if codas consist of randomly scattered S-S conversions, as is commonly assumed. Given the assumption that coda duration is linearly proportional to seismic moment, a magnitude scaling relationship based on coda duration may be written

$$M_{coda} = a \log_{10}(d) + c, \quad (34)$$

where a and c are constants for a particular region, and d is coda duration (Crosson, 1972). We measure coda duration as the time from first motion until peak-to-peak amplitude decays to twice background level.

Duration magnitude scales are usually derived for a large number of hypocenters distributed throughout a source volume. The observed phases for any given earthquake

are received at a subset of the total number of stations. In such a case, the actual coda duration used in equation (34) is an ensemble average of observed coda duration at a number of stations. Ensemble averaging is used to permit a magnitude calibration that is appropriate for use with a typical, but not fixed, set of observed phases. Some estimation variance is necessarily associated with the variation of actual travel paths and scattering features encountered.

For this study, coda duration from only one station (DTN) is used to calibrate the magnitude scale. The coefficients we derive are only applicable to estimation of magnitudes given an observed coda duration at DTN, and an Akutan source location. Thirty-eight events with M_L estimated by the National Earthquake Information Center (NEIC) are used to derive the coefficients a and c in equation (34). A comparison of calibrated duration magnitude estimates with reported M_L and m_b is shown in Figure 34. The best fit values for a and c are 2.36 and 0.4, respectively. Considerable scatter is apparent, as is a consistent offset between M_L and m_b . We adopt M_L as the standard magnitude estimate for this area, primarily because more well determined NEIC estimates of M_L are available than m_b estimates.

Given these magnitude estimates, we calculate seismic moments using the equation

$$M_w = \frac{1}{\beta} [\log_{10}(m_0) - 9], \text{ where } M_w \text{ is moment magnitude, } m_0 \text{ is seismic moment, and } \beta$$

is a function of rupture geometry (Somette et al., 1996). We assume $\beta = 3/2$, which is appropriate for circular rupture (Somette et al., 1996). It has been previously reported that M_L is roughly equivalent to M_w in the range of 4 to 6.5 (Bakun, 1984; Boore and

Joyner, 1982; Katsumata, 1983; Jackson, 1984). We proceed by assuming that our derived $M_{coda} \cong M_L \cong M_w$, and estimate calibrated duration magnitudes for the initial two episodes (Figure 35). Cumulative seismic moment for the initial two episodes is also shown (Figure 36).

The distribution of magnitudes as a function of time within the episodes is strikingly different (Figure 37). During the first episode, 78% of the energy was released in the first six hours ($2.5 \text{ E}17 \text{ J}$), and 99% was released after 22 hours. The greatest additional contribution came in conjunction with several events occurring about twenty hours into the swarm. Energy release during the second episode was much more uniform than the first, with a nearly constant release rate over any hourly period for the first thirteen hours. During the second episode, 78% ($3.77 \text{ E}17 \text{ J}$) the energy was released in the first thirteen hours, and 99% was released after 18 hours. The total seismic energy released during the first two episodes is estimated at $8.0\text{E}17 \text{ Joules}$, which is larger than a single earthquake of $M_w = 5.9$.

Cumulative moment release has been previously used as a criterion for indicating hazardous eruption potential (Yokoyama, 1988). The alarm threshold proposed by Yokoyama(1988) is $10\text{E}10 - 10\text{E}11 \text{ Joules}$, which was exceeded at Akutan by seven orders of magnitude. Akutan provides a case-in-point that energy alone cannot be used as a direct indicator of imminent eruption hazards. The location of strain concentration with respect to probable vent areas must also be considered.

It seems likely that the energetic episode seismicity at Akutan was associated with a magmatic intrusion, given the Akutan environment, the quantity of energy released, and

the sustained event counts over several hours. If the intrusion assumption is accepted, then the cumulative seismic moment can be used to estimate the total volume added to a region (McGarr, 1976; Yokoyama, 1988; Abe 1992). McGarr's relationship is written $M_{cum} = K_0 \mu |\Delta V|$, where M_{cum} is the total seismic moment, $K_0 \cong 1$, μ is the rigidity, and ΔV is the total volume of material added to a region. The relationship was demonstrated to be valid for deformation caused by water in mines, the Matsushiro earthquake swarms, and the Galapagos caldera collapse (McGarr, 1976). If we assume that $\mu = 3.0 \times 10^9 \text{ Nm}^{-2}$, the estimated total change in volume during the Akutan swarm is approximately 0.27 km^3 . The actual value of μ is unknown, and might reasonably be expected to range up to $6.0 \times 10^9 \text{ Nm}^{-2}$, yielding a lower bound on the magmatic injection volume of 0.13 km^3 . The volume change during the most energetic forty hours of seismicity would be associated with a mean magma flow rate during this period of approximately $900\text{-}1800 \text{ m}^3 \text{ s}^{-1}$. The total volume of material would correspond to a VEI 3 eruption (Simkin et al., 1981). Our volumetrically based VEI estimate is conservative because vassiculated tephra volume measurements were used to establish the VEI/volume correspondence. The volume in place, which we calculate, would contain very few vassicles, and would certainly expand during upward transport and eruption.

The distribution of earthquakes as a function of magnitude is often found to satisfy the Gutenberg-Richter frequency-magnitude relationship (Wyss, 1973). The relationship is a power law described by

$$\log_{10}(N) = a - b M, \quad (35)$$

where N is the cumulative number of earthquakes of magnitude greater than or equal to M , and the coefficients a and b are constants within a finite seismogenic region (Ishimoto and Ida, 1939; Gutenberg and Richter, 1954). The value of b is related to the fractal-dimensionality of the dominant rupture process in a volume (Sornette et al., 1991). Rupture dimensionality is a function of material properties and the value of b can be used as a qualitative indicator of material properties in the seismic source volume. Throughout much of the Earth's crust, b is approximately equal to unity (Morgan et al., 1988). Material properties associated with elevated b -values include large material heterogeneity (Mogi, 1962), high geothermal gradients (Warren and Latham, 1970), and elevated shear or effective stress levels (Scholz, 1968; Wyss, 1973; Urbancic et al., 1992).

A variety of methods exist to compute b , given a sequence of earthquakes and magnitude or moment estimates (Aki, 1965; Rundle, 1989; Utsu, 1992; Sornette et al., 1996). Of the proposed methods, the method of rank-ordered statistics performs most robustly. It is particularly appropriate for populations dominated by small events, but which also contain a few extreme events (Sornette et al., 1996). This is the case with the Akutan data, and we use the method of rank-ordered statistics to estimate b .

For Akutan $M_L < 4.5$, we estimate $b = 1.05$ (Figure 38). This value is similar to b -values reported for much of the crust (Morgan et al., 1988), and is somewhat lower than expected for volcanic systems (Wiemer and McNutt, 1997; Wyss et al., 1997). For Akutan $M_L > 4.5$, a notable deviation from the expected distribution occurs. This observation, taken at face value, suggests a different rupture process for the larger events. Either the dominant rupture geometry might be different, so that the error in assuming β

$= 3/2$ is not constant across all magnitudes, or the fractal dimension of the rupture process might be larger. Variation in b as a function of magnitude has been reported for shallow transform earthquakes (Pacheco et al., 1992), and is presumed to exist for subduction zone earthquakes (Sornette et al., 1996). The variation in b for subduction zones is interpreted as the consequence of the transition from 2-D to 1-D dominated rupture. The b -value kink in transform environments has been associated with the magnitude at which the rupture area is limited by the thickness of the seismogenic zone. We presume that the energetic seismicity at Akutan was associated with magmatic injection. Given this presumption, it might be speculated that the smaller events could be associated with planar dike extension, and the larger events could be associated with new crack formation along multiple axes. Such an interpretation is consistent with the apparently higher b -value for events larger than $M_L = 4.5$, but is extremely speculative.

Decay Sequence Data

The last instrumentally recorded event of the second energetic episode occurred March 15, at 13:09 GMT. No additional instrumentally recorded seismic observations are available between this time and 19:16 GMT, March 22, when the emergency array was activated. Data collected with the emergency array were telemetered to the Alaska Volcano Observatory offices in both Fairbanks and Anchorage for real-time monitoring. After activation and prior to July 20, 1996, 1306 locatable events were recorded. Of these 1306 events, 504 had both S and P phase observations with estimated picking errors

of less than 0.5 second at all station locations. Maximum likelihood epicentral location estimates for this subset of data is shown in Figure 33.

Omori Law p-Values for the Decay Sequence

An aftershock sequence of small earthquakes typically occurs after a significant stress perturbation event. Such an event can be either a single large earthquake, or an energetic swarm. As the earthquake volume approaches a new state of stress equilibrium, the earthquake count decays to a background seismicity rate. We refer to the seismicity during this time interval as a decay sequence.

The Omori Law is an empirical relationship describing the rate of aftershock occurrence as a function of time after the main event Omori (1902a,b). The law seems to be almost universally applicable to source volumes seeking a new equilibrium condition after a stress transient. Utsu's modified Omori relation is given by $n(t) = k (t + c)^{-p}$ (Utsu, 1968), where t is the time since the main shock, $n(t)$ is the number of events per unit time, and k , c , and p are free parameters. Parameter k is dependent on the total number of earthquakes in the sequence, c , depends on the rate of activity in the early part of the sequence, and p describes the asymptotic decay rate at the end of the sequence. The units historically used for $n(t)$ are events per day, which we also use.

Numerical modeling of earthquake sequences indicate that p is a function of the material properties in the hypocentral volume (Mikumo and Miyatake, 1979). In particular, high p -values indicate small relaxation times, and the ability of a material to quickly find a new state of stress equilibrium after a stress transient. High p -values have

been associated with a high level of fracturing, and a positive correlation with high surface heat flow is also reported (Kisslinger and Jones, 1991). The value of p is thought to be relatively independent of fault zone heterogeneity and does not appear to be correlated with b -value (Kisslinger and Jones, 1991). Though energetic swarms and single large events are a consequence of fundamentally different processes, the Omori Law still applies during the stress relaxation phase.

Since p is associated with physical properties of interest, it is a useful parameter to use for comparison of earthquake sequences in different areas. The Omori relationship may also be used during a decay sequence to help identify a passive process, or to flag possible active process renewal and increased hazard levels. If the number of earthquakes observed is consistent with that predicted by the Omori relationship, it may be assumed that the observed seismicity is primarily a consequence of passive stress redistribution. If the expected number of earthquakes deviates strongly above the expected number, the existence of an active stress perturbation process might be postulated.

The Omori law is appropriate for most tectonic areas with essentially a zero rate of background seismicity. In volcanic environments, background seismicity levels may be significant, and we include a background rate of seismicity term in our final equation,

$$n(t) = k (t + c)^{-p} + r_0, \quad (36)$$

where r_0 is the estimated background rate of earthquake production. Maximum likelihood estimates of the Omori parameters are easily derived (Ogata, 1982), given a sufficiently long sequence. The reference time, $t = 0$, was taken as the time of the last

event of the second energetic episode. Data from the first few hours of system operation were discarded, and our p-value analysis begins on the first full day of emergency array operation. The data gap between swarm inception and the beginning of our analysis window prevents accurate determination of all the Omori parameters, but p-value estimation is still possible because p is dominated by the decay rate at the end of the sequence.

During the swarm, the rate of background seismicity was unknown. A value of $r_0 = 0.15$, which was based on data from other Aleutian volcanoes, was used during real-time monitoring. The results reported here are based on an $r_0 = 0.25$, which was verified to be appropriate during the second six-month time interval after the swarm. Variation in r_0 affects the absolute value of p, but does not alter tests for consistency with the Omori relationship.

We find the Akutan decay sequence to be consistent with the Omori relationship, and estimate a p-value of 2.52 ± 0.08 (Figure 39). This suggests that the events of the decay sequence were a passive response to the initial stress transient. The distribution of magnitude as a function of time also shows no anomalous increase of seismic energy release at any time during the decay sequence (Figure 40). The p-value is extremely high, relative to the typically reported range of 1.0-1.5 (Mogi, 1962; Utsu, 1968; Utsu, 1969; Mikumo and Miyatake, 1979). The range of reliable p-values of thirty-nine southern California earthquake decay sequences evaluated by Kisslinger and Jones (1991) is 0.69 - 1.81. None of these earthquakes were associated with an active volcanic system however. It is likely that the temperatures in the source area at Akutan were higher than

locations previously reported. Our estimate of p at Akutan might be used to support the proposed correlation of high p -values with elevated temperature.

The 90+ day duration of the Akutan decay sequence is much longer than average for an aftershock sequence following a single $M_w = 5.9$ event (Reasenbergs and Jones, 1989). This does not mean that the seismicity was related to an ongoing magmatic process, only that the nature of the stress field perturbation at Akutan is likely to be quite different from that associated with a single rupture.

Temporal Stability of the Source Concentration Area

The Omori Law provides a guideline for detecting an anomalous increase in seismicity rate, but does not provide a mechanism for detecting anomalous changes in hypocenter concentration areas. Spatial migration of hypocenter concentration has been interpreted as an indicator of magmatically induced stress-field perturbation (Lahr et al. 1994). Such migration is almost certainly an indicator of an ongoing active process. Shifts in hypocenter concentration area have been reported immediately prior to a number of eruptions (Nishi, 1978; Yokoyama et al., 1981; Power, 1988; McNutt, 1996).

Raw arrival times are the most basic information available in earthquake seismology. At the inception of seismology as science, it was recognized that earthquake source locations could be estimated from arrival times by assuming a velocity model and inverting for hypocentral parameters (Geiger, 1912). The process of inversion, however, maps noise in data space into noise in model space. When the number of observations is large (> 20 independent observations), it is reasonable to expect a fairly high level of

noise cancellation, and an accurate estimate of model parameters. When the number of observations is small, data noise maps directly into noise in model space, often in a poorly understood way. Arrays deployed primarily for volcano hazard monitoring purposes are typically small (4-9 stations), and hypocentral estimates can be very sensitive to data noise. Station gain/loss, variation in monitoring personnel, changes in picking style, and intrinsic arrival onset identification limitations all contribute to data noise for the problem of monitoring source volume spatial stability through time.

The spatial stability of hypocenter concentration areas is often monitored using space-time plots of source locations. Noise sensitivity of the inverse method, as discussed in the preceding paragraph, can result in misleading apparent shifts of source locations on such plots. In addition, the intrinsic hypocenter resolution possible with a small aperture array may cause a large amount of intrinsic scattering of source location estimates. Such intrinsic scattering can both obfuscate the existence of actual concentration areas, and hinder the evaluation of source volume spatial stability. We address this issue by posing two distinct questions: 1) Is the spatial distribution of the primary source volume temporally stationary? and 2) Where is the primary source volume?

The spatial stability of hypocenter concentration volumes may be evaluated using relative arrival times directly, without hypocenter location estimation. The basic idea is to monitor the arrival times of each phase, at each station, relative to the P arrival at the most reliably picked station. Concentrated source areas are identified as frequently occurring relative arrival times. A temporally stationary set of hypocenter concentration volumes produces a homogeneous population of relative arrivals throughout the

observation period. Any shift in source volume can be detected by testing a recent set of phase observations for homogeneity with the population of earlier relative arrivals. A variety of statistical procedures are available for this type of application (Hogg and Craig, 1978). We use the Kolmogorov-Smirnov (Daniel, 1990) test for population homogeneity, which makes no distributional assumptions and is more robust than chi-square tests when applied to small populations. At Akutan, the large number of events allowed test populations to include 20 or more samples and the Kolmogorov-Smirnov test was selected primarily because it is designed for use with continuous data.

Relative arrival time monitoring can be automated, is insensitive to station gain/loss, and does not need the multitude of plots along all axes and manual interpretation required for manual / interpretive monitoring of source volume spatial stability. The capability for alarm criteria to be automated is significant within the context of real-time volcano hazard monitoring (English et al., 1988). Relative arrival time population monitoring is simple, effective, robust, and easily implemented.

Within the first few hours of monitoring the Akutan decay sequence, it became clear that the *P* phase at AK2 was identifiable for almost all events. It was therefore selected as the initial reference arrival phase, and was retained as such throughout the sequence. Relative *P* phase times for stations other than AK2 are shown (Figure 41 - Figure 43). More than 90% of the arrivals at AK3 (co-located with AK1 and AKT) arrived at 0.25 ± 0.1 seconds after the AK2 arrival (Figure 41). The small variance in relative arrival time indicates that either: 1) The source volume is very small; or 2) The source volume is elongated in such a way that the relative arrival times between AK2 and AK3

remain almost constant. Considerably more dispersion in relative arrival times at AK4 exists (Figure 42), with 90% of the arrivals occurring at $0.9 (-0.4, +0.2)$ seconds, relative to the AK2 *P* phase. This also suggests a small source volume, but with some elongation. AK5 is the most informative of the raw relative arrival time sets. Two dominant arrival modes exist, at -0.82 and 0.08 seconds. A third fairly strong mode at -0.48 seconds, is also observed (Figure 43). Together, this set of observations suggests a highly concentrated source volume, somewhat elongated, with two primary source regions.

The null hypothesis (homogeneous populations) could not be rejected with 90% or better confidence for any station/phase combination at any time during the monitoring period. This indicates a lack of significant temporal variation of the source volume location. This analysis is consistent with the Omori law conclusion: The decay sequence was a passive, non-spatially varying response to stresses induced by the initial episodes.

Akutan Velocity Model

The problem of determining earthquake locations, seismic velocities, and station corrections is, in general, coupled. When the number of arrival times per event is considerably larger than the number hypocentral parameters being estimated, either simultaneous or iterative estimation of hypocenters and velocity model parameters is possible (Roecker, 1982; Pavlis, 1992). When the number of stations is small, there is seldom enough independent information to allow well-determined velocity model estimation. This is the case at Akutan.

For the decay sequence seismicity at Akutan, we have at most eight independent observations (P and S arrival times at four locations). A simple linear gradient velocity model overlying a constant velocity basement, with a constant V_p/V_s (P velocity to S velocity ratio), is described by four parameters, and hypocentral estimation requires four. The resulting problem of determining eight parameters given eight observations is at best even determined, given sufficient redundant observations to allow assumed picking error cancellation. Nevertheless, we need to adopt a velocity model that is satisfactory for rough earthquake location purposes. Toward this end, we estimate V_p/V_s using a method that does not require estimates of source location or absolute traveltime. We then assume a velocity model consistent with prior experience in geologically similar neighboring areas, refine the model via trial and error, and invert for perturbations to the simple model. We demonstrate model suitability for earthquake location using residual versus distance plots.

Akutan source location estimates are very sensitive to the assumed V_p/V_s . V_p/V_s estimates are usually made using plots of P versus S phase traveltimes at a number of stations (Wadati plots). This implies a need to estimate source locations and origin times. Since these parameters are very sensitive to the assumed V_p/V_s , we seek a method of V_p/V_s estimation that does not rely on hypocentral parameter estimates.

In regions with constant $V_p/V_s = k$, the ratio may be determined without knowledge of earthquake locations, origin times, or traveltimes. Note that for any raypath, the observed traveltime between points a and b is given by $T_{ab} = \int_a^b \frac{1}{V(\underline{x})} ds$, where the

integration is performed along the raypath, and $V(\mathbf{x})$ is the seismic velocity at location \mathbf{x} .

If the ray is assumed to be valid, raypath geometry is controlled by the eikonal equation

$$\frac{d}{ds} \left(\frac{1}{V} \frac{d\mathbf{x}}{ds} \right) = \nabla \left(\frac{1}{V} \right) \quad (\text{Aki and Richards, 1980}).$$

Note that for a constant V_p/V_s ,

$V_s = k^{-1} V_p$, the constant term cancels and the ray equation for both P and S waves is

identical. This implies that for regions with constant V_p/V_s , both P and S waves follow

the same raypath between points. The traveltimes between S and P phases

can then be written as $T_{ab}^{(s)} = k T_{ab}^{(p)}$, where the traveltimes are between points \mathbf{a} and \mathbf{b} ,

$T_{ab}^{(s)}$ is the S wave traveltime, and $T_{ab}^{(p)}$ is the P wave traveltime. Given a single source,

any two stations, \mathbf{a} and \mathbf{b} , the ratio of observed S arrival time to P arrival time differences

can be written

$$\frac{T_{ab}^{(s)} - T_{ac}^{(s)}}{T_{ab}^{(p)} - T_{ac}^{(p)}} = \frac{k T_{ab}^{(p)} - k T_{ac}^{(p)}}{T_{ab}^{(p)} - T_{ac}^{(p)}} = k \frac{V_p}{V_s}. \quad (37)$$

This means that an estimate of V_p/V_s can be made using any two stations observing both

P and S phases for the same event. No source location, distance, or traveltime

calculations are required, given the assumption of constant V_p/V_s .

We now wish to use the collection of point estimates made based on differential arrival times, together with observational uncertainties, to make a grand estimate of V_p/V_s . This may be accomplished in several ways. The simplest approach is to form the grand estimate from a weighted average of individual estimates from equation (37),

$$\left(\frac{V_p}{V_s}\right)^* = \frac{\sum_i^N \frac{1}{\sigma_i} \left(\frac{V_p}{V_s}\right)_i}{\sum_i^N \frac{1}{\sigma_i}}, \quad (38)$$

where N is the number of differential arrival pairs available, $\left(\frac{V_p}{V_s}\right)_i = \left(\frac{T_{ab}^{(s)} - T_{ac}^{(s)}}{T_{ab}^{(p)} - T_{ac}^{(p)}}\right)_i$, and

σ_i^2 is the variance of $\left(\frac{V_p}{V_s}\right)_i$. The variance actually used in the weighted summation, σ_i^2 ,

is obtained by assuming that the picks are uncorrelated and normally distributed, with mean and standard deviation given by the indicated arrival time and reported observational uncertainty, respectively. Both the numerator and denominator of equation (37) are then distributed as the difference of two normal random variables, which is also a normally distributed random variable (Hogg and Craig, 1978). For any two normally distributed phase observations, T_1 and T_2 , the distribution of $(T_2 - T_1)$ is given by

$$f(t_2 - t_1) = N(t_2 - t_1, \sigma_1^2 + \sigma_2^2). \quad (39)$$

To complete the calculation of σ_i^2 , we note that for two normally distributed random variables, X and Y , distributed as $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$, respectively, the mean and variance of the random variable Y/X is given to first order by:

$$\mu_{\left(\frac{y}{x}\right)} = \frac{\mu_y}{\mu_x} \quad \text{and} \quad \sigma_{\left(\frac{y}{x}\right)}^2 = \frac{\mu_y^2 \sigma_x^2 + 3 \sigma_x^2 \sigma_y^2}{\mu_x^4} + \frac{\sigma_y^2}{\mu_x^2} \quad (40).$$

A simple point estimate of V_p/V_s can then be made using matched pairs of phase observations together with equations (38), (39), and (40). A probability density function,

from which error estimates can be derived, can also be constructed using the individual estimates of V_p/V_s . If there are N paired sets of arrival observations, and the i^{th} estimate of V_p/V_s is denoted k_i , with variance σ_i^2 , then the probability density function for any value of k is given by

$$P\left(\frac{V_p}{V_s}\right) = \frac{\sqrt{\sum_{i=1}^N \sigma_i^{-2}}}{(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{1}{2} \sum_{i=1}^N \sigma_i^{-2} \left(k_i - \frac{V_p}{V_s}\right)^2\right] \quad (41)$$

(Menke, 1984). Error estimates and confidence intervals are easily computed, given a probability density function (Hogg and Craig, 1978).

For the Akutan data, 596 matched pairs of P and S picks with estimated picking error of less than 0.1 second are used to estimate $V_p/V_s = 1.75 \pm 0.1$. A scatter-plot of differential arrival times shows no trend toward higher or lower values as a function of differential arrival time (Figure 44). We adopt the value of $V_p/V_s = 1.75$, and do not consider it as a variable during the remainder of rough velocity model development.

Given V_p/V_s a series of linear gradient over a half-space model were tried manually with a small subset of well-picked events. This resulted in a model given by

$$\begin{aligned} V_p(z) &= 3.41 \text{ km s}^{-1} - 0.37 \text{ s}^{-1} z & (z \geq -7 \text{ km}) \\ V_p(z) &= 6 \text{ km s}^{-1} & (z < -7 \text{ km}) \end{aligned}$$

where z = distance above sea, and $z = 0$ is at sea level above. An extra parameter at $z = -20 \text{ km}$ was added to the model, and a linearized inversion for all model parameters other than V_p/V_s was performed. The poorly resolved inverse produced effectively zero

perturbation for the upper parameters, but found $V_p(z = -20 \text{ km}) = 6.3 \text{ km s}^{-1}$. This model is used for the remainder of our analysis.

As noted, the velocity model we use is poorly resolved, and an indicator of its suitability for earthquake location is needed. The variation of traveltime residuals as a function of distance from the observing station may be used as an indicator of velocity model adequacy for earthquake location purposes (Rowlett and Forsyth, 1984). Akutan P and S traveltime residuals for arrivals with estimated picking errors of less than or equal to 0.1 and 0.5 seconds, respectively, are shown in Figure 45. No systematic variation of residual with distance is observed. We conclude that the velocity model is adequate for rough earthquake location purposes.

Hypocenter Estimation: Issues, Requirements, and Methods

Hypocenter determination is limited by picking accuracy, the number of phases available, the aperture of the picked phase locations with respect to hypocenter location, and the accuracy of forward traveltime computation. When using models with strong velocity contrasts, layers, topography, or a small array, the location problem becomes non-linear and an appropriate inversion algorithm must be selected to avoid local minima (Shen et al., 1984). Local minima may also appear when station aperture is limited ("mirroring"). The apparent spatial distribution of located hypocenters after locating a group of events is a function of the actual spatial distribution, mapped through the intrinsic ambiguity function associated with the inversion process.

Physical interpretation of the spatial distribution of hypocenters requires accurate assessment of the intrinsic ambiguity of the location process, or prior knowledge that the location uncertainty is small relative to the desired scale of physical interpretation.

At Akutan, a small number of stations are available and the aperture is limited. The velocity structure is poorly constrained, and traveltimes calculations that rely on smooth-earth models are inconsistent with the shallow hypocenters observed and the rugged topography of the island. Nevertheless, we seek to accurately assess the hypocenter distribution and interpret the result within the context of active physical processes. To address these demands, we need to compute traveltimes through arbitrarily complex velocity models (with topography), avoid local minima, and accurately assess the expected spatial distribution of hypocenters intrinsic to the inverse method used.

Many methods of hypocentral estimation exist (Geiger, 1912; Bolt, 1960; Evernden, 1969; Lee and Lahr, 1975; Hermann, 1979; Lahr, 1980; Pavlis and Booker, 1980; Tarantola and Valette, 1982; Thurber, 1985; Hirata and Matsu'ura, 1987; Klein, 1988; Prugger and Gendzwill, 1988; Nelson and Vidale, 1990; Abers and Roecker, 1991; Moser et al., 1992), some of which address the needs outlined here better than others. The methods fall into three primary categories: 1) Global search methods; 2) Downhill methods; 3) Joint inverse methods. All methods require calculation of theoretical traveltimes from a specified source location to a number of specified receiver locations, and use some form of prediction error, or cost, function. Global search methods (Nelson and Vidale, 1990; Zhou, 1992; Moser et al., 1992) typically search a schedule of grid-nodes over a selected *a priori* volume for the node with the smallest prediction error.

These methods are robust, and find global minima, but require many traveltime calculations. They are only feasible if an extremely high-speed traveltime calculation method is available. Downhill methods are the most widely used, and attempt to iteratively proceed from some hypothesized location toward one with a smaller prediction error. The downhill strategy is most often linearized using first order partial derivatives of prediction error with respect to location (Lee and Lahr, 1975; Hermann, 1979), but step-search methods (Prugger and Gendzwill, 1988), and inclusion of second order terms (Thurber, 1985) can also be used. All downhill methods may descend into local minima, if “mirroring” effects are present, and thus lack the robustness of global search methods. Joint inverse methods may be useful if the objective is to simultaneously estimate both velocity structure and source parameters for a large number of hypocenters (Pavlis and Booker, 1980; Roecker, 1982). One notable exception to the primary categories listed above is the method introduced by Tarantola and Valette (1982). In their approach, a likelihood density function is computed throughout an *a priori* volume. It is then possible to integrate the density function to obtain expected, rather than most-likely, estimates of source location. In both the methods of Tarantola and Valette (1982) and Hirata and Matsu’ura (1987), explicit reference to origin time is not required to compute relative source location likelihood.

The Akutan emergency array, as indicated, is small and has a very limited aperture. “Mirroring” problems became evident during initial location estimation efforts. For this reason, a global search method is needed. We employ a grid-search method with nonlinear conjugate-gradient (Gill et al., 1981) downhill refinement from the best node.

A high-speed traveltime calculation algorithm, which handles raypaths in the presence of rugged topography, is used (Chapter 4, this thesis). We also experiment with several prediction error criteria to determine the sensitivity of hypocentral location estimates to *a priori* noise distribution assumptions.

Several options exist for the prediction error function. If Gaussian noise is assumed and data errors are uncorrelated, the zero-covariance, weighted L_2 norm cost function is used. It is given by

$$E_{L_2}^2(\mathbf{x}) = \frac{\sum_j^n \sigma_j^{-2} (\tau_j - \tau_0 - h_j)^2}{\sum_j^n \sigma_j^{-2}} \quad (42)$$

where τ_j is the observed arrival time of the j^{th} phase, τ_0 is the origin time, h_j is the theoretical traveltime from location \mathbf{x} to the station location of the j^{th} phase observation, and σ_j is the estimated standard deviation of the phase identification error. $E(\mathbf{x})$ is often referred to as the weighted RMS error.

The optimal estimated origin time using the L_2 norm is

$$\tau_0^{(L_2)} = \frac{\sum_j^n \sigma_j^{-2} (\tau_j - h_j)}{\sum_j^n \sigma_j^{-2}} \quad (43)$$

The associated conditional joint density function is

$$f_{L_2}(\mathbf{x}) \sim \exp\left(\frac{-E_{L_2}^2(\mathbf{x})}{2}\right). \quad (44)$$

When only a small number of observations are available, L_2 norm solutions may be very sensitive to noise. A more robust approach is to assume exponentially distributed noise (Draper and Smith, 1981), which implies use of the L_1 norm cost function (Menke, 1984). Use of the L_1 norm cost function should result in more reliable hypocenter estimates (Prugger and Gendzwill, 1988; Nelson and Vidale, 1990). The L_1 norm cost function is given by

$$E_{L_1}(\mathbf{x}) = \frac{\sum_j^n \sigma_j^{-1} |\tau_j - \tau_0 - h_j|}{\sum_j^n \sigma_j^{-1}}. \quad (45)$$

The most likely origin time using the L_1 norm is

$$\tau_0^{(L_1)} = \frac{\sum_j^n \sigma_j^{-1} (\tau_j - h_j)}{\sum_j^n \sigma_j^{-1}}$$

and the density function is

$$f_{L_1}(\mathbf{x}) \sim \exp(-E_{L_1}(\mathbf{x})). \quad (46)$$

Two potential deficiencies exist with the standard L_1 and L_2 norm approaches: 1) An infinite number of equally acceptable solutions may exist for poorly picked events; and 2) Modeling errors are not considered. Both of these problems have been addressed (Tarantola and Valette, 1982; Moser et al., 1992). Their generalized cost function for the case of Gaussian data and noise is given by

$$E_{G_2}^2(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} r_i(\mathbf{x}) r_j(\mathbf{x}) \quad (47)$$

where $r_i(\mathbf{x}) = \tau_i - \frac{\sum_{k=1}^n w_{ik} \tau_k}{\sum_{k=1}^n w_{ik}} - h_i(\mathbf{x}) + \frac{\sum_{k=1}^n w_{ik} h_k(\mathbf{x})}{\sum_{k=1}^n w_{ik}}$, and w_{ij} is the element from the i^{th} row

and j^{th} column of matrix $(\mathbf{C}^{(d)} + \mathbf{C}^{(m)})^{-1}$, $\mathbf{C}^{(d)}$ is the data covariance matrix, and $\mathbf{C}^{(m)}$ is a

modeling error covariance matrix. The terms of the modeling error covariance matrix

may be chosen to have the form $\mathbf{C}_{ij}^{(m)} = \sigma_i^{(m)} \sigma_j^{(m)} \exp\left[-\frac{1}{2} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\Delta^2}\right]$, where $\sigma_i^{(m)}$ is the

standard deviation of modeling error for the i^{th} ray, and Δ is some correlation length

chosen to reflect the spatial correlation of velocity perturbations away from the actual

model.

The density function is given by

$$f_{G_2}(\mathbf{x}) \sim \exp(-E_{G_2}^2(x)). \quad (48)$$

The most likely origin time is given by $\tau_0^{(G_2)}(\mathbf{x}) = \frac{\sum_i \sum_j w_{ij} (\tau_i - h_i(\mathbf{x}))}{\sum_i \sum_j w_{ij}}$, but is not

required in the calculation of the cost and density functions. This approach theoretically accounts for modeling errors and data covariance. Poorly posed problems are naturally stabilized by the modeling error term, without use of explicit *a priori* information about probable source location.

After selecting an appropriate conditional hypocenter location joint density function, such as those given above, error estimates are easily computed. After a normalization

constant κ is selected such that $\kappa \iiint_V f(x) dV = 1$, marginal density functions along any axis or plane can be obtained by integration along any other axis or axes. For example, if, x_1 and x_2 are chosen to be horizontal axes, and depth is associated with x_3 , then the marginal probability function for the depth of a hypocenter is given by

$$f(x_3) = \kappa \iint_V f(x_1, x_2) dx_1 dx_2. \text{ The expected value for the depth of the hypocenter is}$$

$$\text{given by } \langle x_3 \rangle = \kappa \int_{-\infty}^{\infty} x_3 f(x_3) dx_3, \text{ and the variance is given by}$$

$$\text{var}(x_3) = \kappa \int_{-\infty}^{\infty} x_3^2 f(x_3) dx_3 - \langle x_3 \rangle^2. \text{ Epicentral joint density functions are obtained with}$$

only one integration, as

$$f(x_1, x_2) = \kappa \int_{-\infty}^{\infty} f(x_3) dx_3. \quad (49)$$

Error estimates along any axis are obtained in a similar fashion. Practically, the integrations are performed numerically over some *a priori* area or volume, or with limits that are allowed to expand until the contribution to the integral becomes small. An *a priori* density function may be multiplicatively applied to the conditional density functions either before or after to integration. In this case another normalization constant may be required to convert the final density function into a probability density function.

There are times when the need for an accurate error assessment does not justify the full computation of a density function throughout the likely source volume. In this case,

a linearized covariance estimate can be made about the most likely point (Abers, 1996).

The equation is given by

$$\mathbf{C}_m = \left[\mathbf{G}^T (\mathbf{C}_d + \mathbf{C}_t)^{-1} \mathbf{G} + \mathbf{C}_p \right]^{-1}, \quad (50)$$

where

\mathbf{C}_m = covariance of model parameter estimate (source coordinates)

\mathbf{G}^T = partial derivatives of cost function with respect to the source coordinates

\mathbf{C}_d = prior estimate of the covariance of the data

\mathbf{C}_t = prior estimate of the covariance of modeling errors

\mathbf{C}_p = prior estimate of the covariance of the source coordinates (usually zero matrix).

This approximation is primarily useful for making quick approximate error estimates, or when computing weights for tomographic inversions.

Akutan Hypocenter Estimates

Maximum likelihood hypocentral locations of estimates for 504 earthquakes with both *P* and *S* picks at all four station locations are shown in Figure 33. All events shown have maximum estimated picking errors of 0.1 second for *P* phases, and 0.5 seconds for *S* phases. The location estimates are optimized using the Tarantola-Valette L_2 cost function, equation (48). The spatial correlation distance used is very small, so that observations from the co-located stations are averaged, and the other observations are considered uncorrelated. A 1% modeling error was assumed, which is certainly an underestimate of the actual error associated with the velocity model. This results in optimistic (smaller than is probable) error estimates.

The locus of seismicity is observed in an area 6-10 kilometers east, and 2 kilometers south, of the summit (Figure 33). A diffuse east-west scattering of epicenters over a four kilometer distance is observed, as well as a concentrated NE-SW linear trend. An area of detail in this vicinity was selected for further investigation (Figure 33). An epicentral density function was computed for this area using the most likely relative arrival times (Figure 46). The modal arrival pattern used for inversion was selected by first considering only events with P arrival estimated picking errors of less than 0.1 second at both AK2 and AK5. Then events with an AK5-AK2 arrival time of 0.0-0.2 seconds were selected. Empirical relative arrival time density functions for these data are constructed and the most likely arrival times were used for location density function analysis. Location error estimates for the density function were optimistically set at 0.1 seconds for all phases. The intrinsic location uncertainty indicated by the epicentral density function is remarkably coincident with the strong epicentral trend. We conclude that the NE-SW trend of epicenters is due to intrinsic estimation uncertainty, and does not indicate a dike injection along the line.

A similar analysis was performed for arrivals fitting the second modal pattern with AK5-AK2 relative P arrival times of -0.82 ± 0.8 seconds. This arrival pattern is clearly associated with the hypocenters west of the primary concentration area (Figure 49). The epicentral density function is much less directional than that associated with the dominant mode arrivals. We conclude that sources in this region may be distributed through a volume that may be somewhat larger than a point-source, although much of the scatter is

directly attributable to estimation ambiguity. No apparent hypocenter scattering suggesting a source concentration along a fault, conduit, or dike system is apparent.

A cross-section of the density function along line A-A' also indicates a high degree of correlation between the intrinsic location uncertainty and the maximum likelihood hypocenter projections into the plane. The nonlinear character of the function is reminiscent of Rosenbrock's function (Gill et al., 1981), which is often used to demonstrate convergence problems for downhill based optimization methods. This illustrates the need for full global nonlinear source location estimation. Depth is poorly resolved, and the beginning of a "mirroring" effect is observed near the surface (Figure 47). Hypocenter location algorithms that do not account for surface topography are prone to extend such mirrors upwards. Optimal numerical solutions located well above the surface are then possible. Such solutions do not necessarily indicate a 'bad' set of picks, just that the optimization algorithm descended into a non-physical local minimum.

An L_1 norm density function in the same plane displays similar characteristics, but with a smaller region of maximum likelihood. This is expected. The actual error depends upon the validity of the data error distribution assumption, and the accuracy of the phase identification error estimates.

Most of the Akutan seismicity can be attributed to two or three small, closely spaced, sources in the area shown. Errors associated with source location estimation are too large to permit interpretation of a detailed source location geometry. Depth errors for most of the events are larger than 2 kilometers, but almost all the events can be assigned a depth of 0-5 kilometers with 90% confidence. Some of the east-west scattering is judged to be

real, but some is certainly associated with difficult *S* picks at AK1, AK2, and AK4. The NE-SW epicentral trend, as well as the apparent northward deepening of hypocenters along line A-A', is associated only with inversion uncertainty, and does not represent a physical phenomenon.

DISCUSSION

The mechanics of the swarm are still a mystery, though several pieces of evidence are consistent with the following speculative scenario:

- 1) Deformation associated with magmatic intrusion caused the Akutan swarm;
- 2) Approximately $0.13 - 0.27 \text{ km}^3$ of magmatic material was injected;
- 3) Active deformation occurred during a time period of less than forty hours;
- 4) The depth of intrusion was 2-4 kilometers below sea-level;
- 5) Decay sequence seismicity was associated with stress-field relaxation in the vicinity of the intrusion, approximately 10 km from the caldera.

The Akutan seismic swarm was unusual in its rapid onset and intensity. The total energy released and the time frame associated with the first two episodes is consistent with a magma flow rate of $900\text{-}1800 \text{ m}^3 \text{ s}^{-1}$.

The intrusion probably occurred in a small region near the decay sequence hypocenter concentration. The region is approximately 10 kilometers from any recent vent area. The large numbers of earthquakes in the source concentration volume, and the lack of earthquakes elsewhere, is not consistent with the alternative hypotheses of island wide fault activation, or inflation near a recent vent.

The cumulative energy released exceeded Yokoyama's eruption alarm threshold by 7 orders of magnitude. Some surface manifestation of the swarm would be expected had the intrusion been associated with an active vent area. No eruption occurred. No tremor, LP events, or other seismic signals indicative of a "breathing" system were observed. No evidence of elevated surface heat flow or venting was observed during a July, 1996, Alaska Volcano Observatory field expedition. Some scattered ground-cracks were observed, in a NW-SE trend at several points on the island, but as indicated, decay sequence epicenters do not support the idea of large-scale through-going fault activation.

A linear feature on the decay sequence epicentral plots, which might be interpreted as evidence of significant dike injection along a NE-SW trending plane, is attributable to intrinsic location uncertainty. While some E-W scatter of sources is apparently real, the data are consistent with the idea of a spatially small source region.

Decay sequence p-values are consistent with the modified Omori Law, indicating that decay sequence seismicity was probably due to stress field relaxation after the initial episodes of energetic seismicity. Relative arrival time subsets are homogeneous with respect to the global population throughout the decay sequence, indicating a spatially stationary source volume. No additional pulses of energetic seismicity occurred. Both these pieces of evidence support the idea of stress-field relaxation in the vicinity of intrusion, and an absence of additional stress-field perturbation caused by ongoing magmatic processes.

The most likely relative arrival pattern of the decay sequence was not observed during a time period of 6-10 months after the swarm. If the primary decay sequence

source volume had been associated with a region simply prone to a high level of seismicity given a regional stress-field perturbation, it would be expected to maintain its status as an area of seismic source concentration. This is not the case. The source volume appears to have “healed” in a period of less than six-months.

CONCLUSIONS

An energetic earthquake swarm and extended decay sequence occurred during the spring of 1996 at Akutan Volcano, Alaska. The cumulative seismic moment for the initial episodes is estimated to exceed 8.0×10^{17} Joules. More felt events occurred during these two episodes than in any other swarm in the Global Volcano Earthquake Swarm Database (Benoit and McNutt, 1996). No eruption, or other surface manifestation of surface venting, occurred.

The Gutenberg-Richter b-value at Akutan is approximately 1.05. This value is typical for much of the earth’s crust, and is somewhat low for a volcanic region. A b-value kink toward a higher value for events larger than $M_L = 4.5$ is suggested by the magnitude/rank-order plot, perhaps indicating a different rupture process for the larger events.

The time distribution of events during the decay sequence is consistent with the modified Omori Law ($p = 2.5$), suggesting that decay sequence seismicity was associated with passive stress field relaxation in a high-temperature and/or highly fractured hypocentral volume.

Decay sequence seismicity is concentrated in an area approximately eight kilometers east of the caldera, at a depth of less than four kilometers. A NE-SW trending linear feature apparent on epicenter plots is attributable to intrinsic estimation error. Relative arrival times for events during the decay sequence indicate that the primary source volume was stationary throughout the monitoring period.

We speculate that the initial swarm was associated with two pulses of magmatic intrusion near the hypocenter concentration areas. A volumetric inflation of $0.13 - 0.27 \text{ km}^3$ in the region of observed hypocenter concentration is consistent with the observed seismicity. Hypocenter estimation errors are too large to support a detailed interpretation of intrusion geometry.

ACKNOWLEDGEMENTS

Projects were funded in part by NSF grants EAR-94-05471, EAR-95-06379, the Alaska Volcano Observatory, and United States Geological Survey grant 14-08-0001-A0574 administered through the Alaska Volcano Observatory. Scholarships were provided by the Alaska Geophysical Society and Exxon through the SEG Scholarship Foundation .

Computational resources used for this investigation were provided by the University of Alaska Fairbanks Geophysical Institute (Seismology Laboratory), the Alaska Volcano Observatory, the U.S. Geological Survey (Anchorage, AK), and Applied Digital Technologies.

John Power, John Paskievitch, and Steve Estes deserve thanks for installation of the Akutan emergency array. Steve McNutt and John Power led the seismological and hazard monitoring effort. Art Jolly, John Benoit, and the Fairbanks AVO community in general performed most of effort required for daily monitoring. John Eichelberger and Terry Keith made it possible for the primary author to visit Akutan during the summer field program following the swarm. John Power, Art Jolly, John Benoit, Pete Stelling, Game McGimsey, and Chris Waythomas provided valuable commentary concerning the concepts presented here. Thomas P. Miller provided a non-trivial component of the motivation for pursuing this investigation. The opinions presented here are those of the authors alone and do not necessarily reflect the opinion of the Alaska Volcano Observatory or the University of Alaska Fairbanks Geophysical Institute.

REFERENCES

- Abe, K., 1992 Seismicity of the caldera-making eruption of Mount Katmai, Alaska in 1912, Bull. Seismol. Soc. Am: **82**, 175-191.
- Abers, G. A., 1994, Three-dimensional inversion of regional *P* and *S* arrival times in the East Aleutions and sources of subduction zone gravity highs: J. Geophys. Res., **89**, 6079-9094.
- Abers, G. A., and Roecker, S., 1991, Deep structure of an arc-continent collision: Earthquake relocation and joint inversions for upper mantle *P* and *S* wave velocities beneath Papua New Guinea: J. Geophys. Res., **96**, 6379-6401.
- Aki, K., and Richards, P. G., 1980, Quantitative seismology theory and methods, W.H. Freeman and Company, 1980.

- Bakun, W.H., 1984, Seismic moments, local magnitudes, and coda-duration magnitudes for earthquakes in central California: *Bull. Seismol. Soc. Am.*, **74**, 439-458.
- Benoit, J. P., and McNutt, S. R., 1996, Global volcanic earthquake swarm database and preliminary analysis of volcanic earthquake swarm duration: *Annali de Geofisica*, **39**, 221-229.
- Bolt, B. A., 1960, The revision of earthquake epicenters, focal depths and origin times, using a high-speed computer, *Geophysical Journal*, **3**, 433-440.
- Boore, D. M., and Joyner W. B., 1982, The empirical prediction of ground motion: *Bull. Seismol. Soc. Am.*, **74**, S43-S60.
- Crosson, R. S., 1972, Small Earthquakes, structure, and tectonics of the Puget Sound Region: *Bull. Seismol. Soc. Am.*, **71**, 1133-1171.
- Daniel, W. D., 1990, Applied nonparametric statistics: PWS-Kent Publishing.
- Draper, N. R. and Smith, H., 1981, Applied regression: Wiley, New York.
- English, T. T., Johnson, C. E., Koyanagi, R. Y., and Largo, A., 1988, Automated real-time systems for managing volcanic and seismic crises at the USGS Hawaiian Volcano Observatory: *Seism. Res. Lett.*, **59**, 37.
- Evernden, J. F., 1969, Precision of epicenters obtained by small numbers of world-wide stations: *Bull. Seismol. Soc. Amer.*, **59**, 1365-1398.
- Geiger, L., 1912, Probability method for the determination of earthquake epicenters from arrival times only, *Bull. St. Louis, Univ.*, **8**, 60-71.
- Gill, P. E., Murray, W., and Wright, M. H., 1981, Practical optimization: Academic Press.
- Gutenberg, B., and C. F. Richter, Seismicity of the Earth and associated phenomenon, 2nd edition, Princeton University Press, 1954.

- Herrmann, R. B., 1979, FASTHYPO-A hypocenter location program: Earthquake notes, **50**, 25-37.
- Hirata, N., and Matsu'ura, M., 1987, Maximum likelihood estimation of hypocenter with origin time eliminated using nonlinear inversion technique, Phys. Earth Planet. Inter., **47**, 50-61.
- Hogg, R. V., and Craig, A. T., 1978, Introduction to Mathematical Statistics, 4th edition: Macmillan Publishing Co., Inc.
- Ishimoto, M., and Ida, K., 1939, Observations of earthquakes registered with the microseismograph constructed recently: Bull. Earthq. Res. Inst., **17**, 443-478.
- Jackson, S. M., 1994, Magnitudes of earthquakes associated with basalt dike intrusion for use in INEL seismic hazards evaluations: Idaho National Engineering Laboratory.
- Katsumata, A., 1983, Comparison between the JMA magnitude and the moment magnitude and estimation of the attenuation function for the maximum displacement amplitude on the basis of moment magnitude: in International workshop on strong motion data, pp. 127-144, Port Harbour Research Institute of Japan and United States Geological Survey, USA.
- Kisslinger, C., and Jones, L. M., 1991, Properties of aftershock sequences in Southern California: J. Geophys. Res., **97**, 11,947-11,958.
- Klein, F. W., 1988, User's guide to HYPOINVERSE, a program program for VAX computers to solve for earthquake locations, U.S. Geol. Surv. Open File Report 88.
- Lahr, J. C., 1980, HYPOELLIPSE/MULTICS: A computer program for determining local earthquake hypocentral parameters, magnitude, and first motion pattern: U.S. Geol. Surv. Open-File Report 80-59.

- Lahr, J. C., Chouet, B. A., Stephens, C. D., Power, J. A., and Page, B. A., 1994, Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989-1990 eruptions at Redoubt Volcano, Alaska: *Journ. Volcan. Geotherm. Res.*, **62**, 137-151.
- Lee, W. H. K., and Lahr, J. C., 1975, HYPO71 a computer program for determining hypocenter, magnitude, and first-motion pattern of earthquakes: U.S. Geol. Surv. Open-File Report 88.
- McGarr, A., 1976, Seismic moments and volume change, *J. Geophys. Res.*, **81**: 1487-1494.
- McNutt, S. R., 1996, Seismic monitoring and eruption forecasting of volcanoes: A review of the state-of-the-art and case histories. In *Monitoring and mitigation of volcano hazards*, R. Scarpa and R. I. Tilling eds. Springer-Verlag Berlin, 99-146.
- Menke, W., 1984, *Geophysical data analysis: Discrete inverse theory*: Academic Press Inc.
- Mikumo, T., and Miyatake, M., 1979, Earthquake sequences on a frictional fault model with non-uniform strengths and relaxation times *Geoph. Jour. Roy. Astron. Soc.*, **59**, 497-522.
- Miller, T. P., McGimsey, R. G., Richter, D. H., Riehle, J. R., Nye, C. J., Yount, M. E., Dumoulin, J. A , 1997, Catalog of the historically active volcanoes of Alaska, U.S..G.S Open File Report, in press.
- Mogi, K, 1962, on the time distribution of aftershocks accompanying the recent major earthquakes in and around Japan: *Bull. Earthq. Res. Inst., Tokyo Univ.*, **40**, 107-124.

- Mogi, K., 1963, Some discussions on aftershocks, foreshocks, and earthquake swarms--
The fracture of a semi-infinite body caused by an inner stress origin and its
relation to the earthquake phenomena, 3: Bulletin of the Earthquake Research
Institute, v. **41**, p. 615-658.
- Morgan, F. D., Wadge, G., Latchman, J. Aspinall, W. P., Hudson, D., and Samstag, F.,
1988, The earthquake hazard alert of September 1982 in southern Tobago: Bull.
Seismol. Soc. Amer., **78**, 1550-1562.
- Moser, T. J., van Eck, T., and Nolet, G., 1992, Hypocenter determination in strongly
heterogeneous earth models using the shortest path method: J. Geophys. Res., **97**,
6563-6572.
- Nelson, G. D., and Vidale, J. E., 1990, Earthquake locations by 3-D finite-difference
travel times: Bull. Seismol. Soc. Am., **80**, 395-410.
- Nishi, K., 1978, On the focal mechanism of volcanic earthquakes in Sakarajima volcano:
Ann. Rep. Dis. Prev. Res. Inst., Kyoto Univ., **21**, 145-152.
- Ogata, Y., 1982, Estimation of the parameters in the modified Omori formula for
aftershock frequencies by the maximum likelihood procedure: J. Phys. Earth, **31**,
115-124.
- Omori, F., 1902 a, Macro seismic measurement in Tokyo (1) Publ. Earthq. Invest.
Comm.: **10**, 1-102.
- Omori, F., 1902 b, Macro seismic measurement in Tokyo (2) Publ. Earthq. Invest.
Comm.: **10**, 1-77.
- Pacheco, J. F., Scholz, C. H., and Sykes, L. R., 1992, Changes in frequency-size
relationship from small to large earthquakes: Nature, **355**, 71-73.

- Pavlis, G. L., and Booker, J. R., 1980, The mixed discrete-continuous inverse problem: application to the simultaneous determination of hypocenters and velocity structure: *J. Geophys. Res.*, **85**, 4801-4810.
- Pavlis, G. L., 1992, Appraising relative earthquake location errors: *Bull. Seismol. Soc. Am.*, **82**, 836-859.
- Power, J. A., 1988, Seismicity associated with the 1986 eruption of Augustine Volcano, Alaska: M.S. thesis, University of Alaska Fairbanks.
- Prugger, A. F., and Gendzwill, D. J., 1988, Microearthquake location: a nonlinear approach that makes use of a simplex stepping procedure, *Bull. Seismol. Soc. Am.*, **78**, 799-815.
- Reasenber, P. A., and Jones, L., 1989, Earhtquake hazard after a mainshock: *Sci.*, **243**, 1173-1176.
- Roecker, S. W., 1982, Velocity structure of the Pamir-Hindu Kush region: Possible evidence of a subducted crust: *J. Geophys. Res.*, **87**, 945-959.
- Rowlett, H., and Forsyth, D. W., 1984, Recent faulting and microearthquakes at the intersection of the Verma Fracture Zone and the Mid-Atlantic Ridge: *J. Geophys. Res.*, **89**, 6079-9094.
- Shen, Y., Forsyth, D. W., Condor J., and Dorman, L. M., 1984, Investigation of microearthquake activity following an intraplate teleseismic swarm on the west flank of the Southern East Pacific Rise: *J. Geophys. Res.*, **89**, 6079-9094.
- Simkin, T., Siebert, L., McLelland, L., Bridge, D., Newhall, C., and Latter, J. H., 1981, *Volcanoes of the world*: Smithsonian Institution, Hutchinson Ross Publ. Co.
- Sornette, D., Vanneste, C., and Sornette, A., 1991, Dispersion of b-values in Gutenberg-Richter law as a consequence of a proposed fractal nature of continental faulting: *Geophys. Res. Lett.*, **18**, 897-900.

- Sornette, D., Knopoff, L., Kagan, Y. Y., and Vanneste, C., 1996, Rank-ordering statistics of extreme events: Application to the distribution of large earthquakes: *J. Geophys. Res.*, **101**, 13,883-13,893.
- Tarantola, A., and Valette, B., 1982, Inverse Problems = quest for information, *J. Geophys.*, **50**, 159-170.
- Thurber, C. H., 1985, Nonlinear earthquake location: Theory and examples: *Bull. Seismol. Soc. Amer.*, **75**, 779-790.
- Utsu, T., 1968, A note on the statistical nature of energy and strain release in earthquake sequences: *J. Fac. Sci. Hokkaido Univ., Series VII, Geophysics*, **3**, 49-64.
- Utsu, T., 1969, Aftershocks and earthquake statistics (1), *J. Fac. Sci. Hokkaido Univ., Series VII*, **3**, 130-195.
- Utsu, T., 1992, A method of determining the value of b in a formula $n = a - bM$ showing the magnitude frequency for earthquakes: *Pageoph.*, **139**, 449-462.
- Yokoyama, I., Yamashita, H., Watanabe, H., and Okada, H., 1981, Geophysical characteristics of dacite volcanism – The 1977-1978 eruption of Usu volcano: *Journ. Volcan. Geotherm. Res.*, **9**, 335-358.
- Yokoyama, I., 1988, Seismic energy release from volcanoes: *Bull. Volcan.*, **50**, 1-13.
- Wiemer, S., and McNutt, S. R., 1997, Variations in the frequency magnitude distribution with depth in two volcanic areas: Mount St. Helens, Washington, and Mount Spurr, Alaska, *Geoph. Res. Lett.*, **24**, 189-192.
- Wyss, M., 1973, Towards a physical understanding of the earthquake frequency distribution: *Geoph. Jour. Roy. Astron. Soc.*, **31**, 341-359.
- Wyss, M., Shimazaki, K., and Wiemer, S., 1997, Mapping active magma chambers by b -values beneath off-Ito Volcano, Japan, (in press).

Zhou, H. W., 1992, Rapid and accurate earthquake location by a 3-D network ray-tracing method: Eos Trans. AGU, **73**, 197.

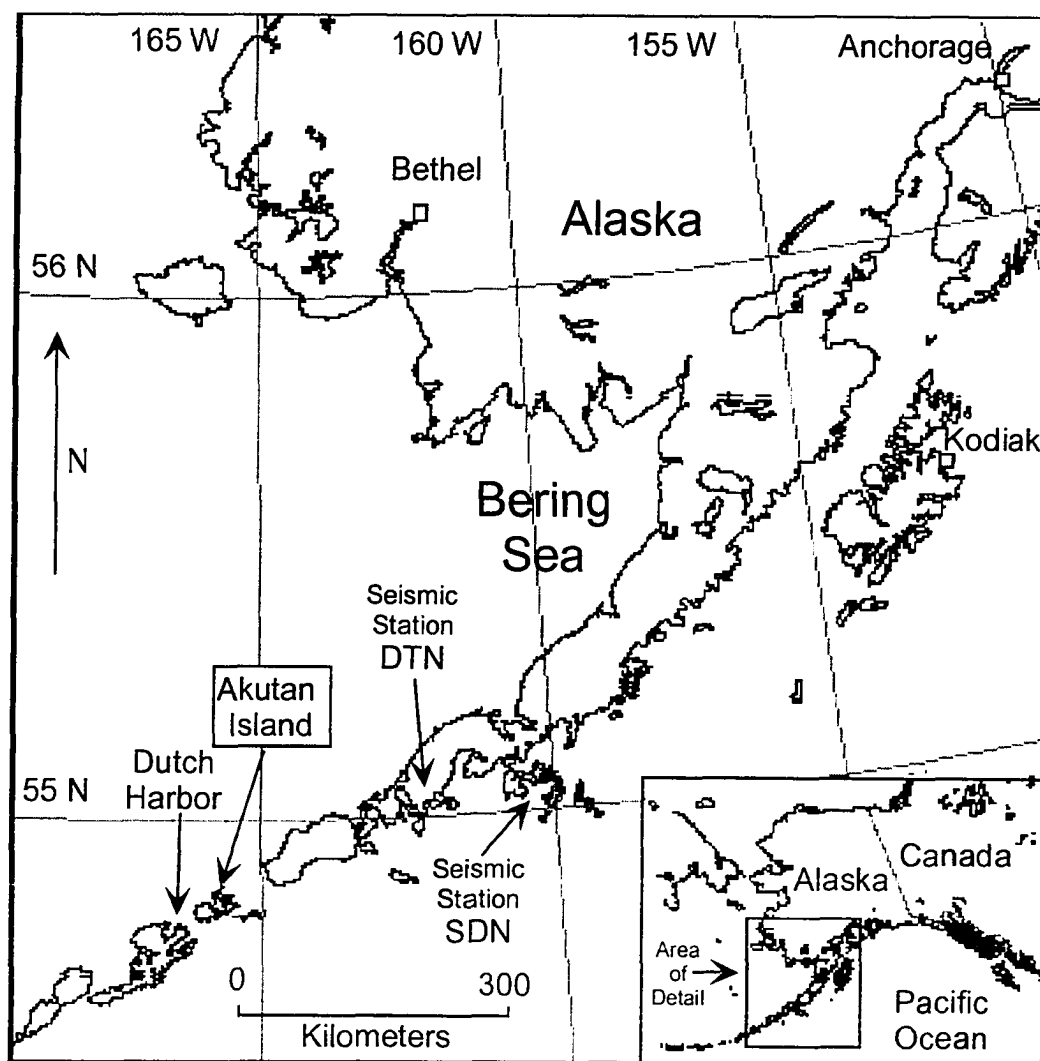


Figure 31: Location of Akutan Island and station DTN.

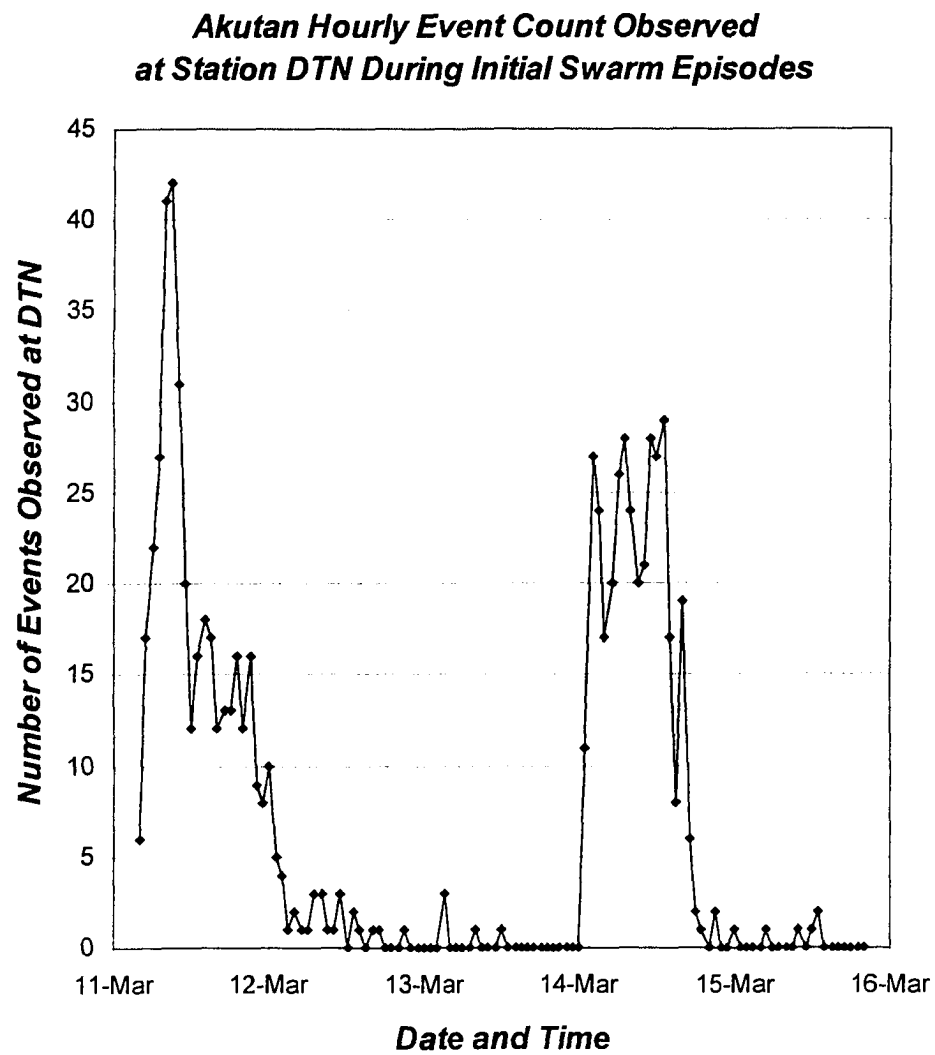


Figure 32: Akutan event counts per hour observed at station DTN.

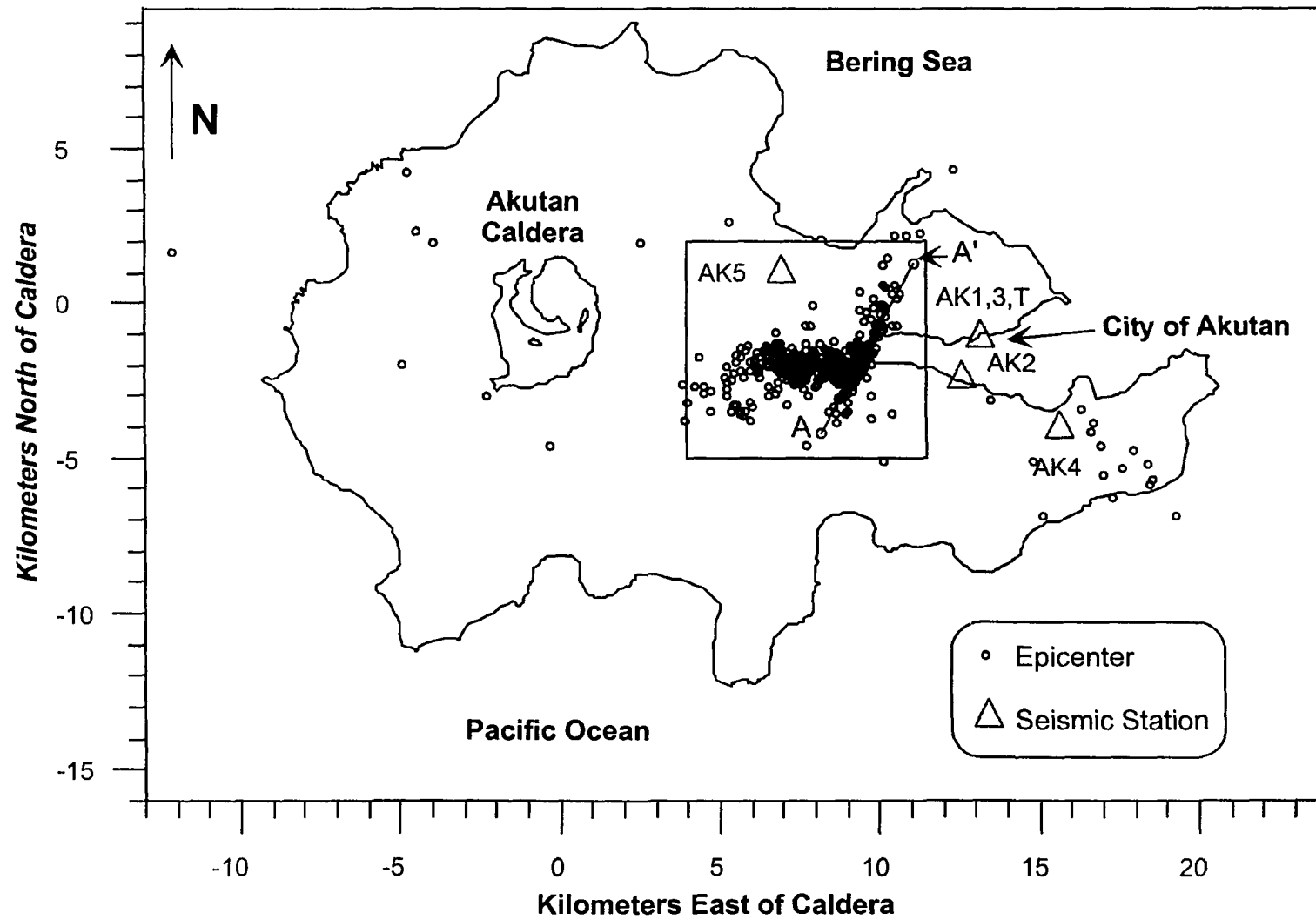


Figure 33: Map of Akutan Island showing location of seismic stations, caldera, city, and decay sequence epicenters.

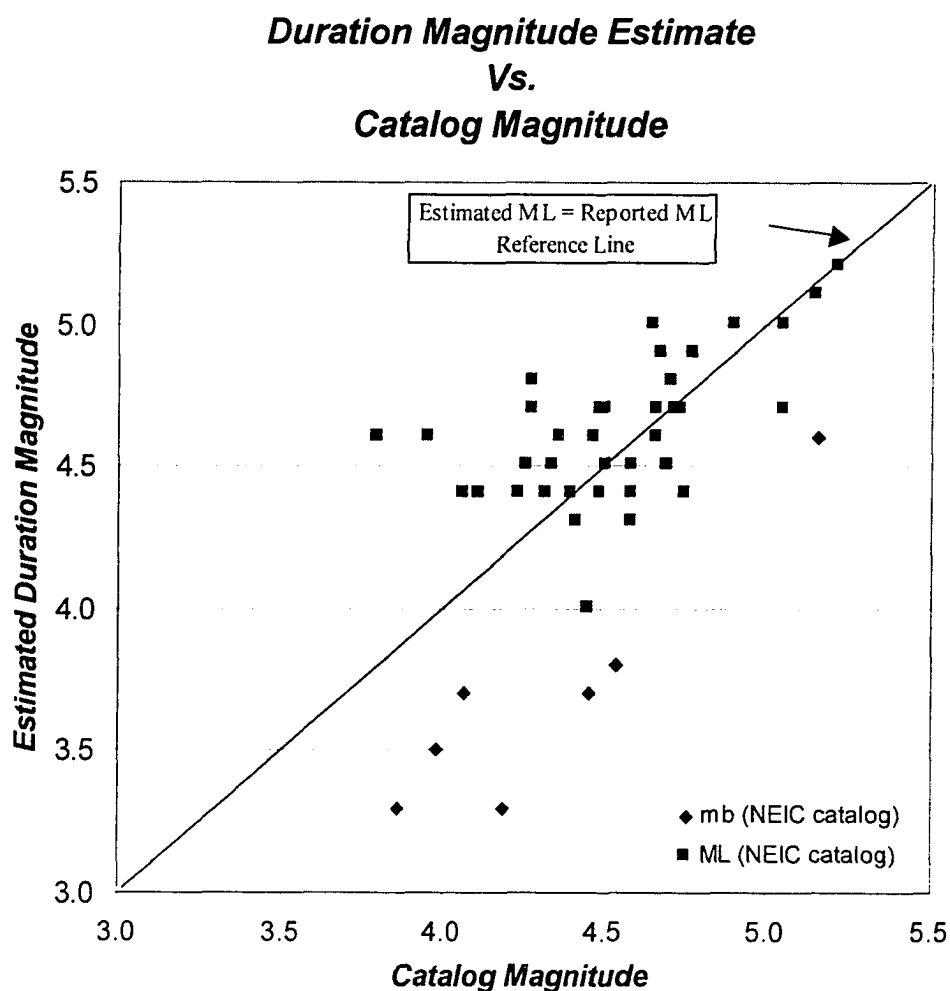


Figure 34: Estimated duration magnitude versus AEIC or NEIC catalog magnitude. The duration magnitude is calculated using $M = 2.36 \log_{10}(\text{duration}) - 0.4$. The coefficients were derived by matching predicted amplitudes to M_L estimates from the AEIC catalog. The Akutan data m_b estimates reported by NEIC appear to differ from AEIC reported M_L values by an additive magnitude factor.

Earthquake Magnitude Vs. Date/Time

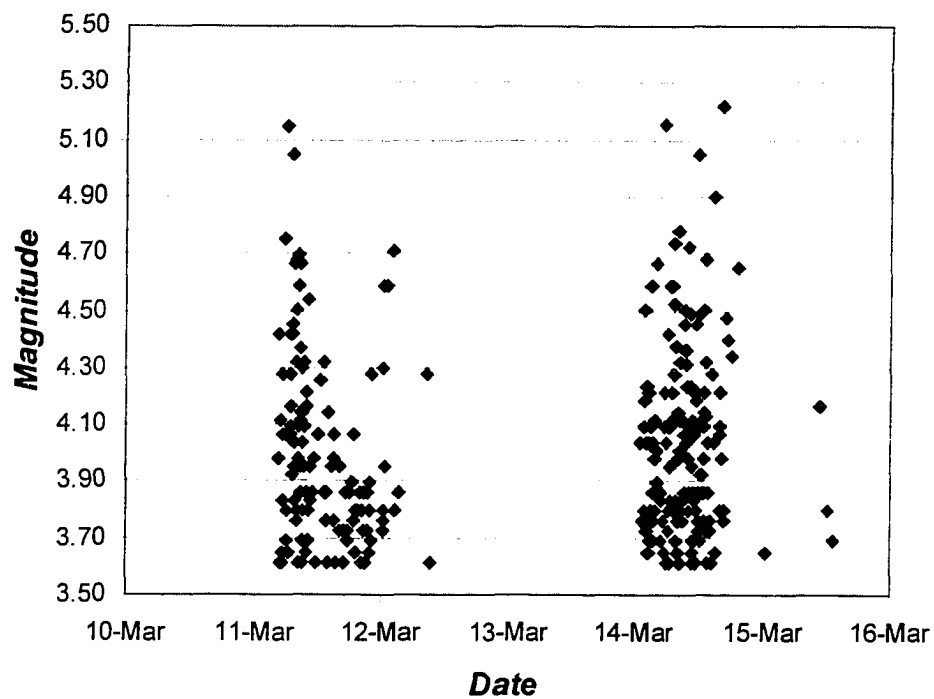


Figure 35: Earthquake magnitude versus time for the first two episodes of the Akutan swarm.

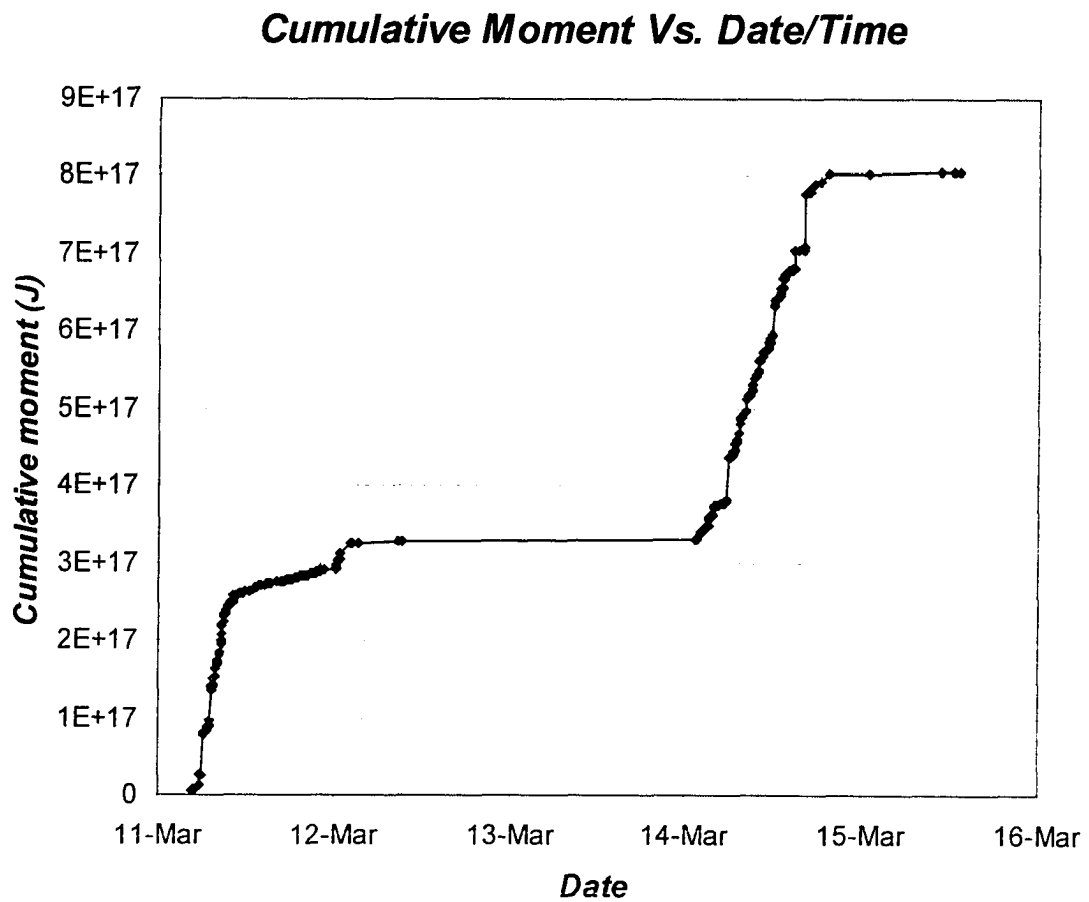


Figure 36: Cumulative moment versus date for the first two episodes of seismicity.

Comparison of Cumulative Energy Release Since Beginning of Episode

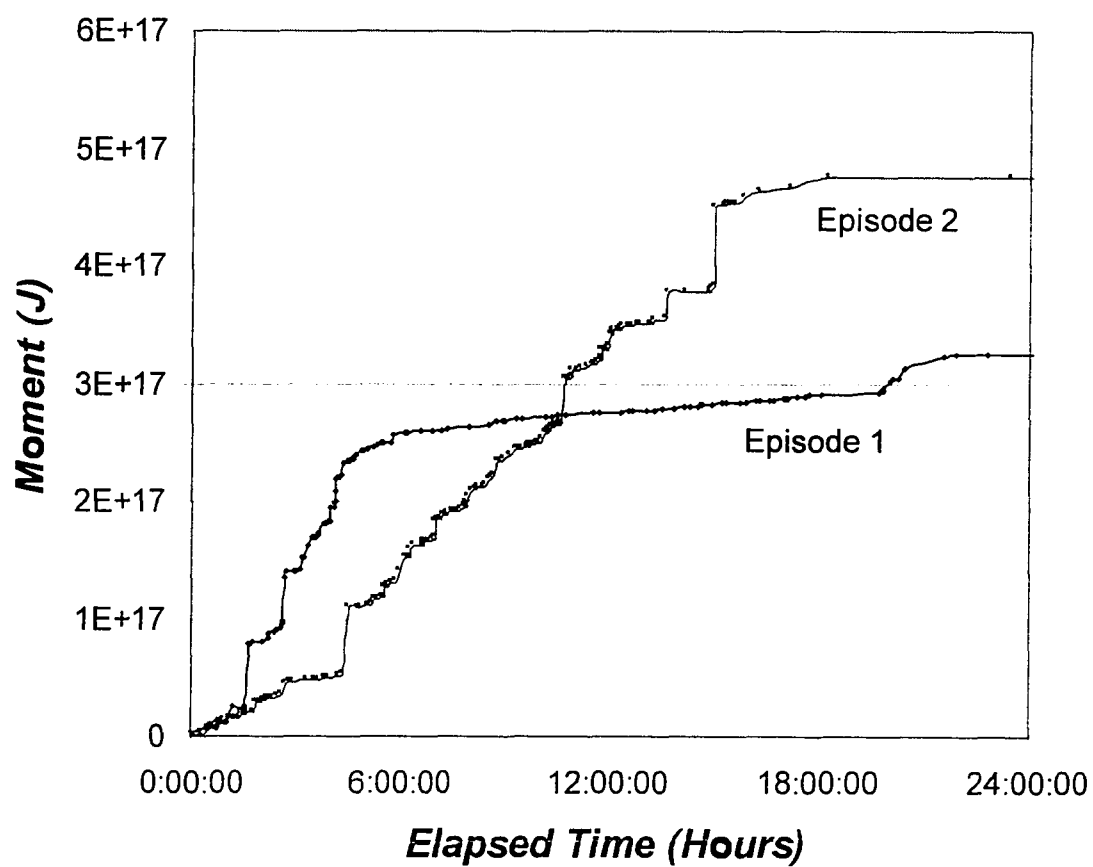


Figure 37: Cumulative energy release versus elapsed time after episode inception for the first two episodes.

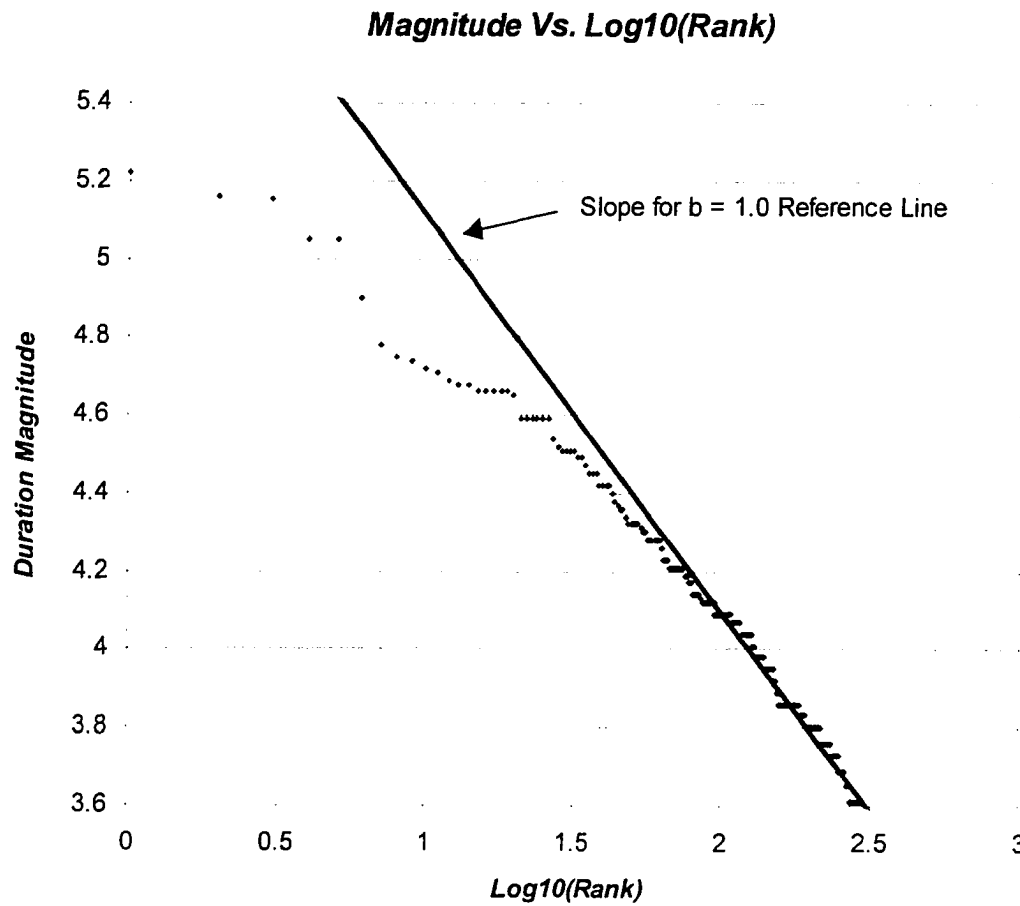


Figure 38: Magnitude versus Log10 (rank).

Below $M_L = 4.5$, a b-value of 1 - 1.05 is consistent with the smaller events. Considerable deviation from this value is apparent for the larger earthquakes.

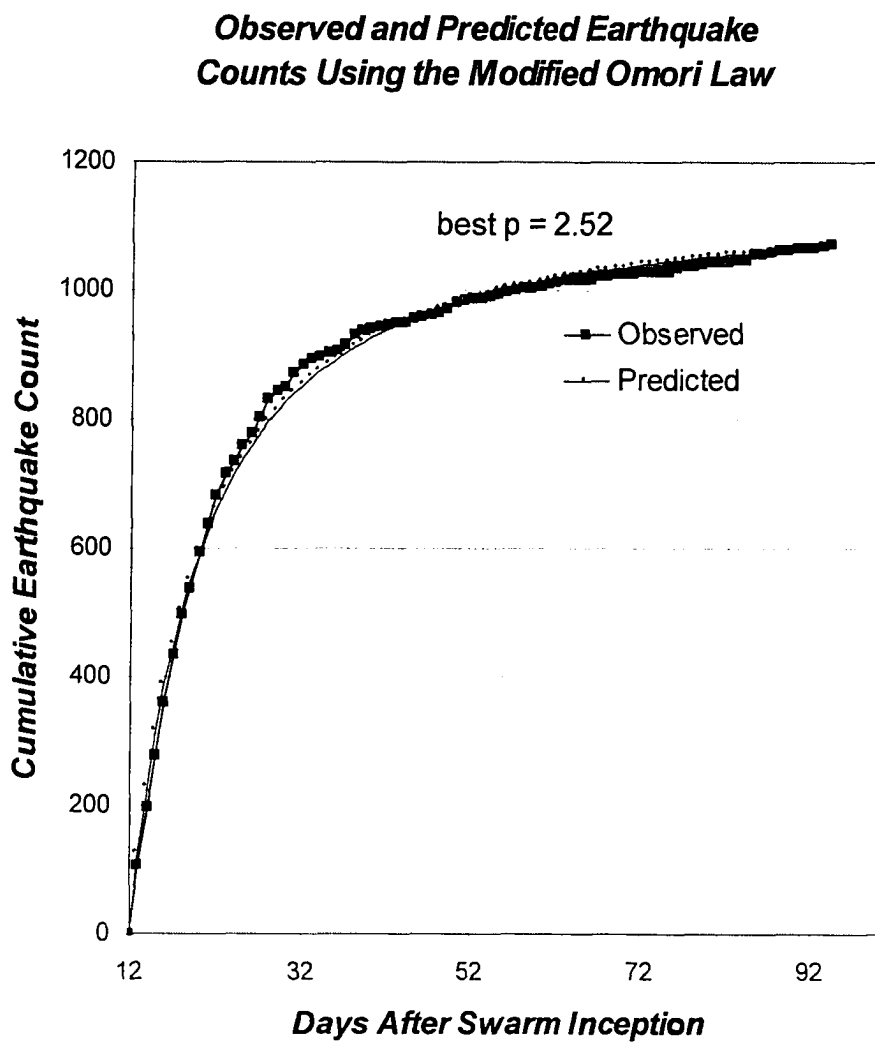


Figure 39: Comparison of observed earthquake counts with those predicted by the modified Omori Law.

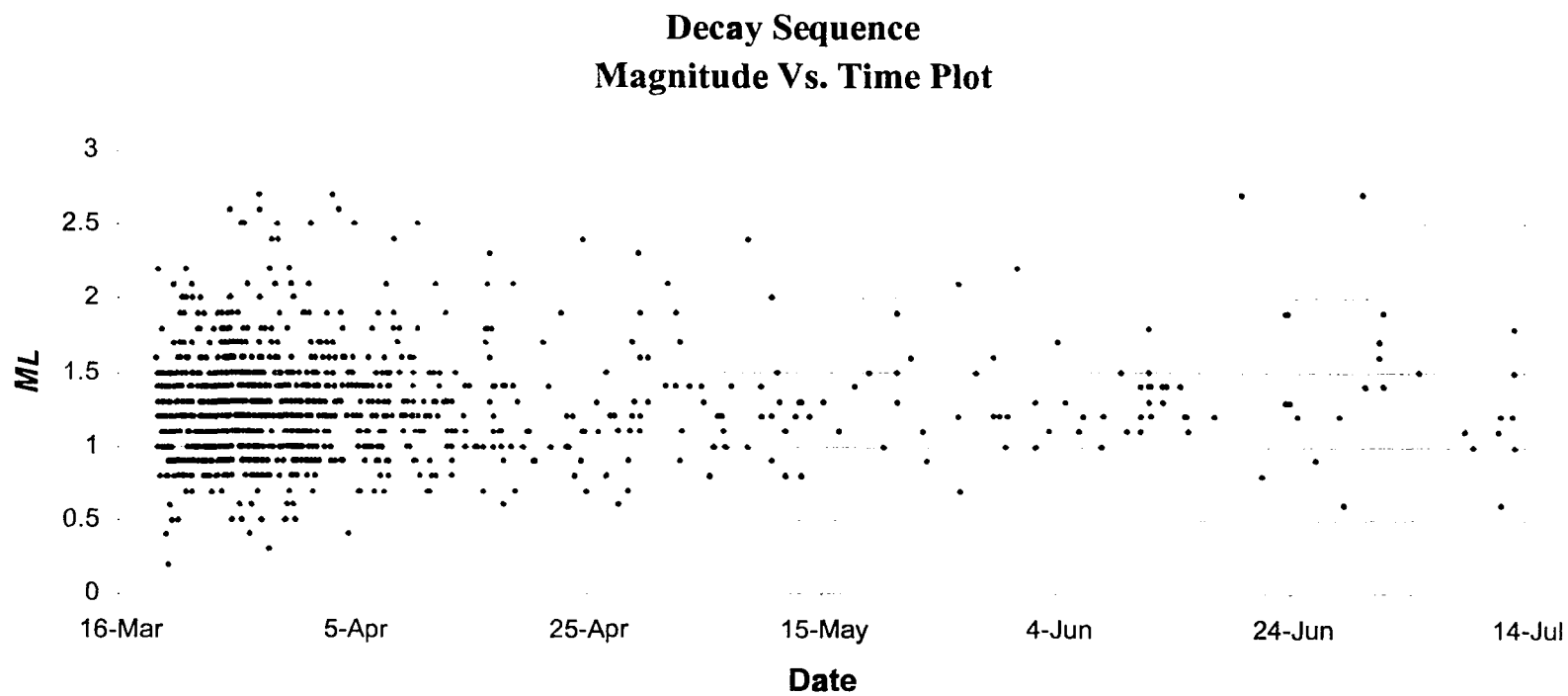


Figure 40: Decay sequence magnitude versus time.

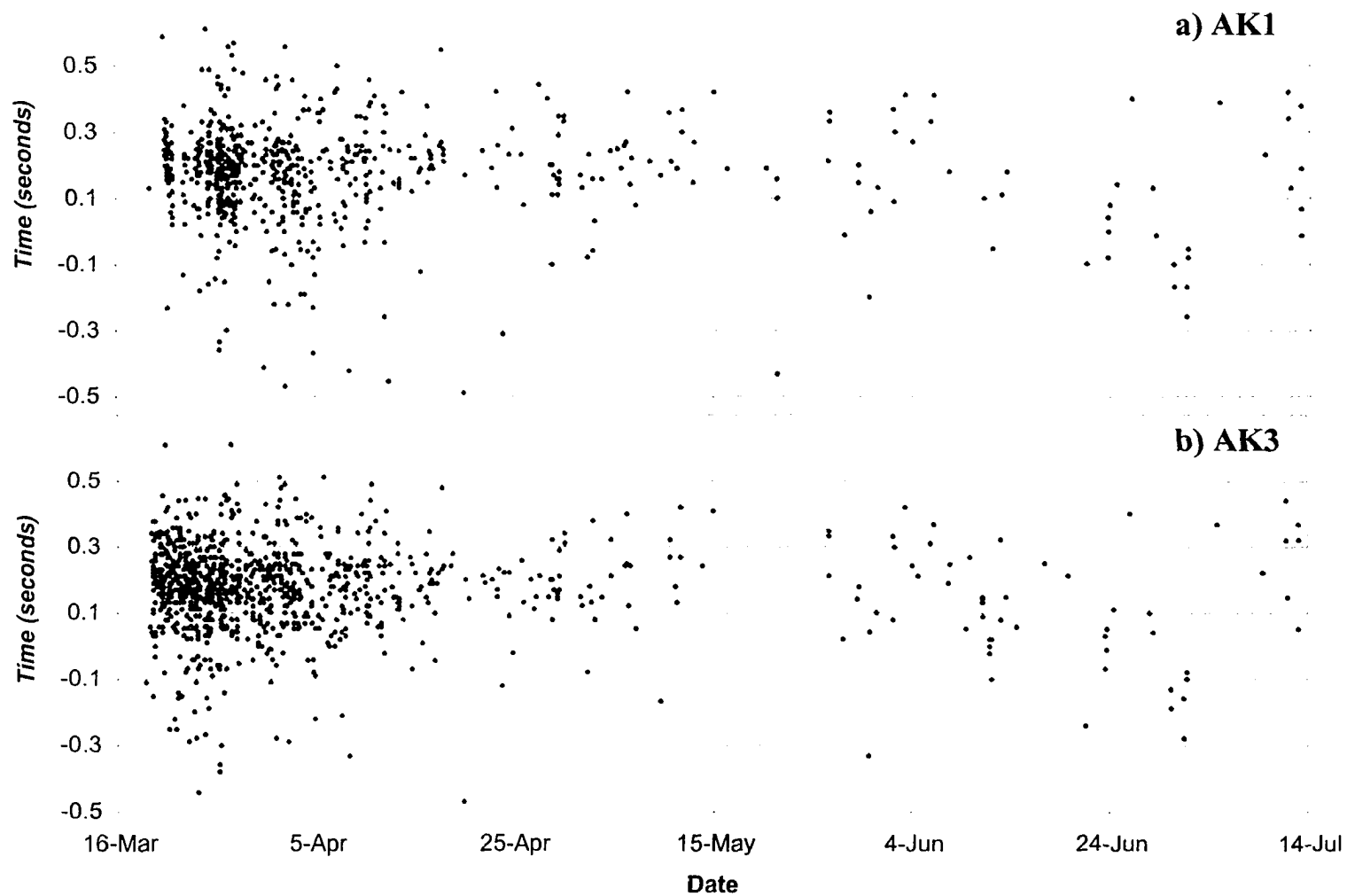


Figure 41: P-wave arrival time at AK1 and AK3 relative to AK2 versus date and time of observation. a) Ak1; b) Ak3: Both stations are co-located. AKT is also colocated, but is not depicted here.

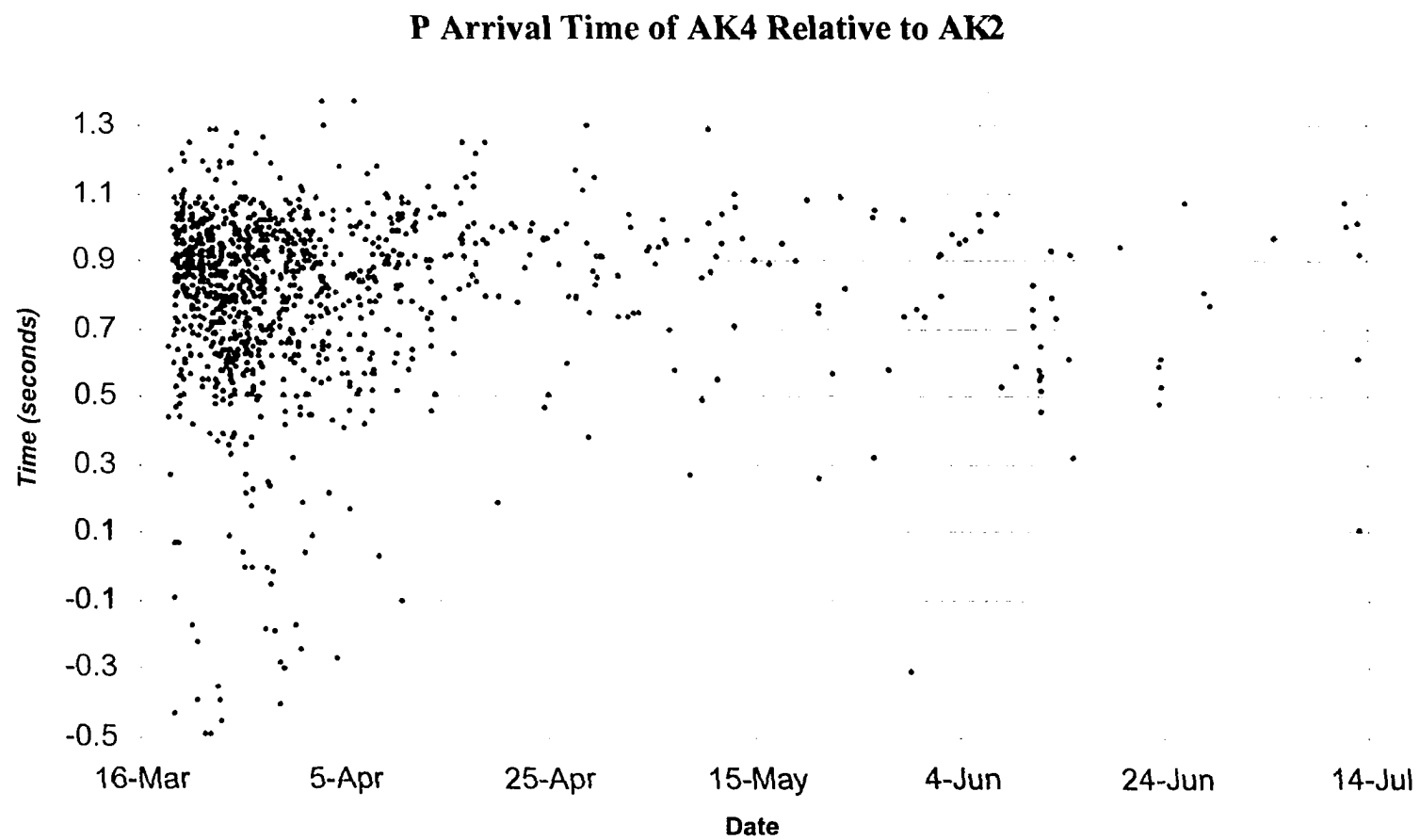


Figure 42: P-wave arrival time at AK4 relative to AK2 versus date and time of observation.

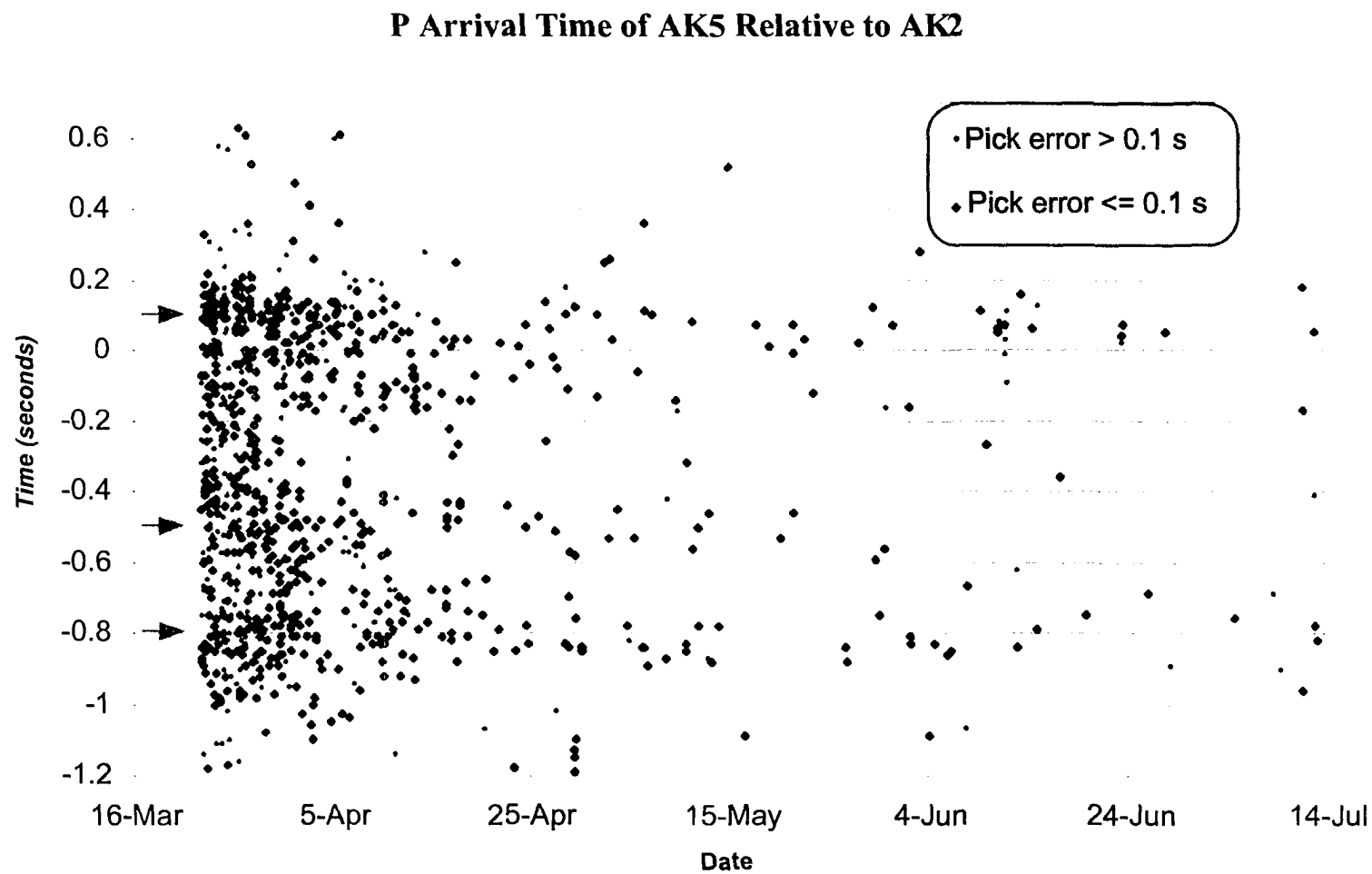


Figure 43: P-wave arrival time at AK5 relative to AK2 versus date and time of observation. Dominant arrival modes are indicated with arrows.

Delta P Vs. Delta S

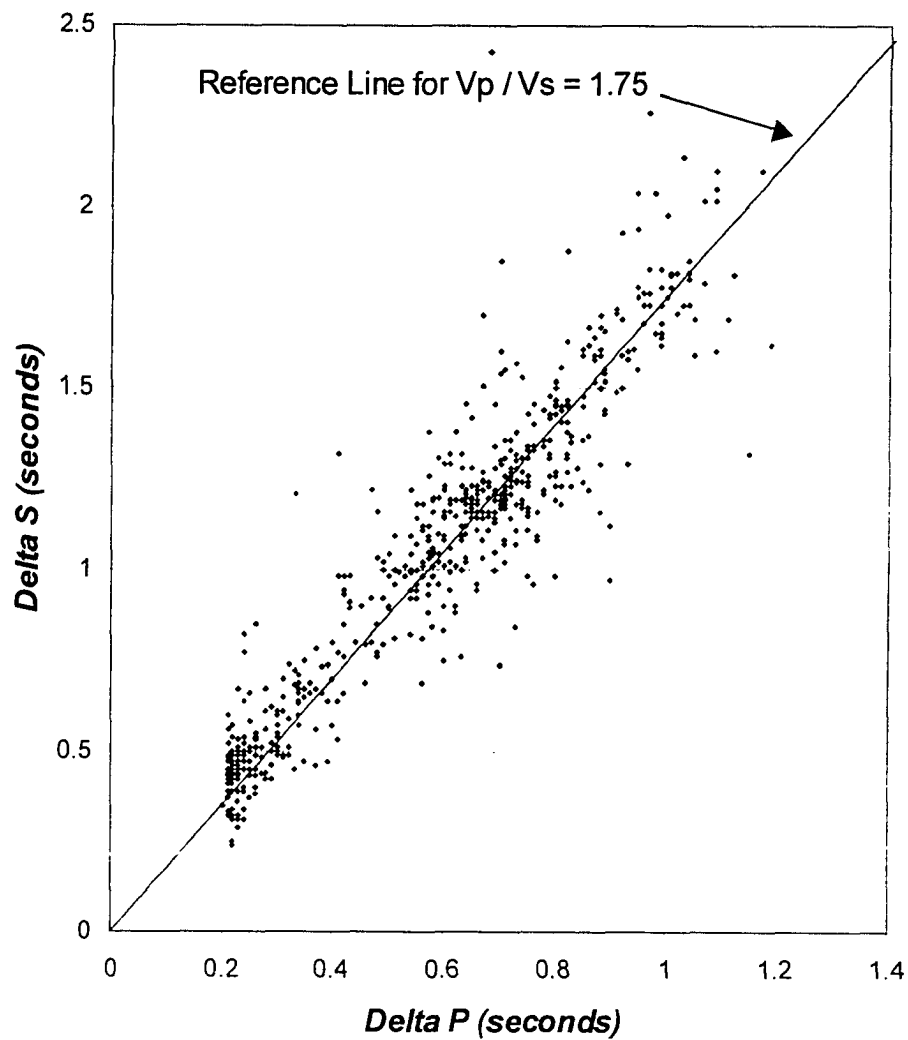


Figure 44: Ratio of *P*-wave to *S*-wave traveltime differences used for estimating V_p/V_s . Data from 546 matched pairs of arrivals with estimated picking errors of 0.1 second or less are shown.

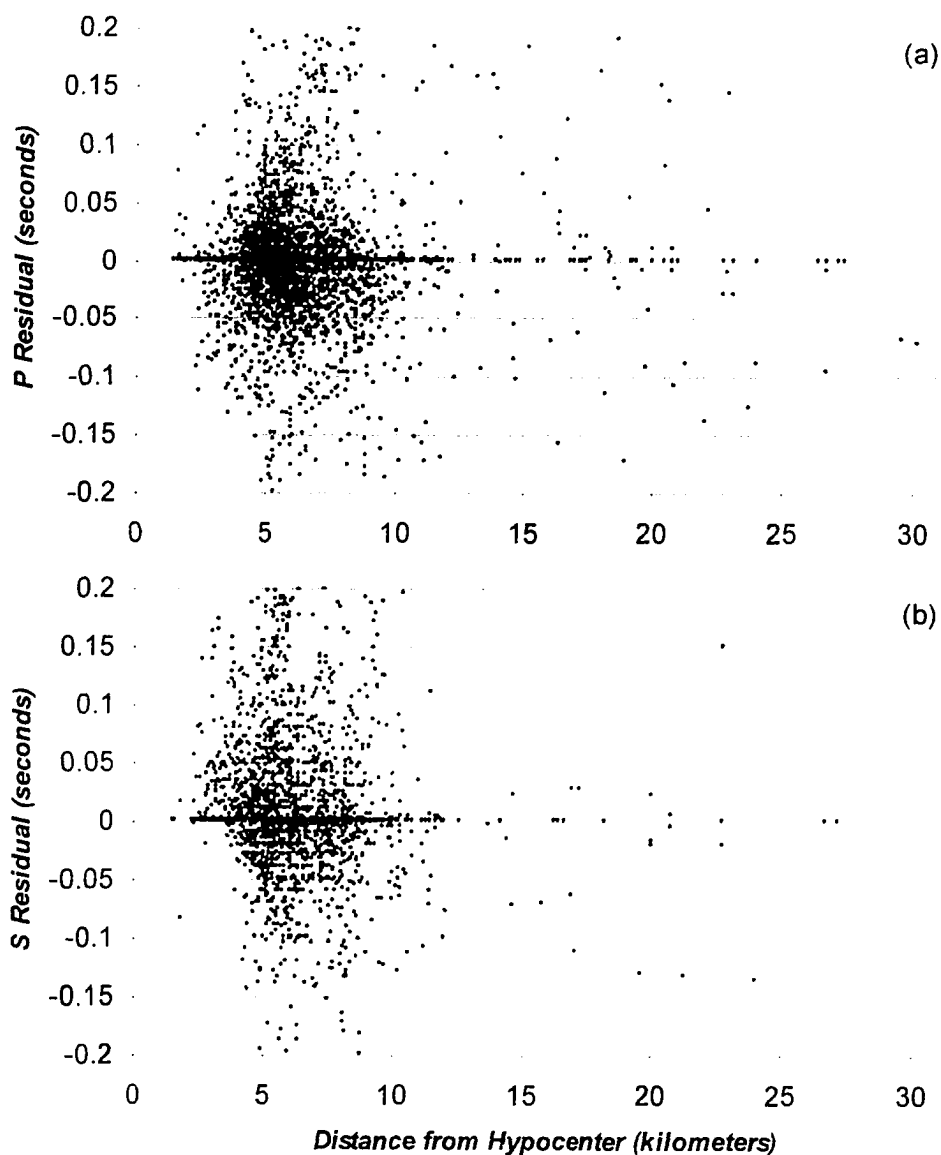


Figure 45: Traveltime residuals versus distance from hypocenter. No systematic variation of residual as a function of distance is observed, suggesting velocity model adequacy for earthquake location purposes. a) 4441 *P* residuals with estimated picking errors of less than or equal to 0.1 second. b) 2830 *S* residuals with estimated picking errors of less than or equal to 0.1 second.

Epicentral Density Function and 432 Hypocenters in Detail Area

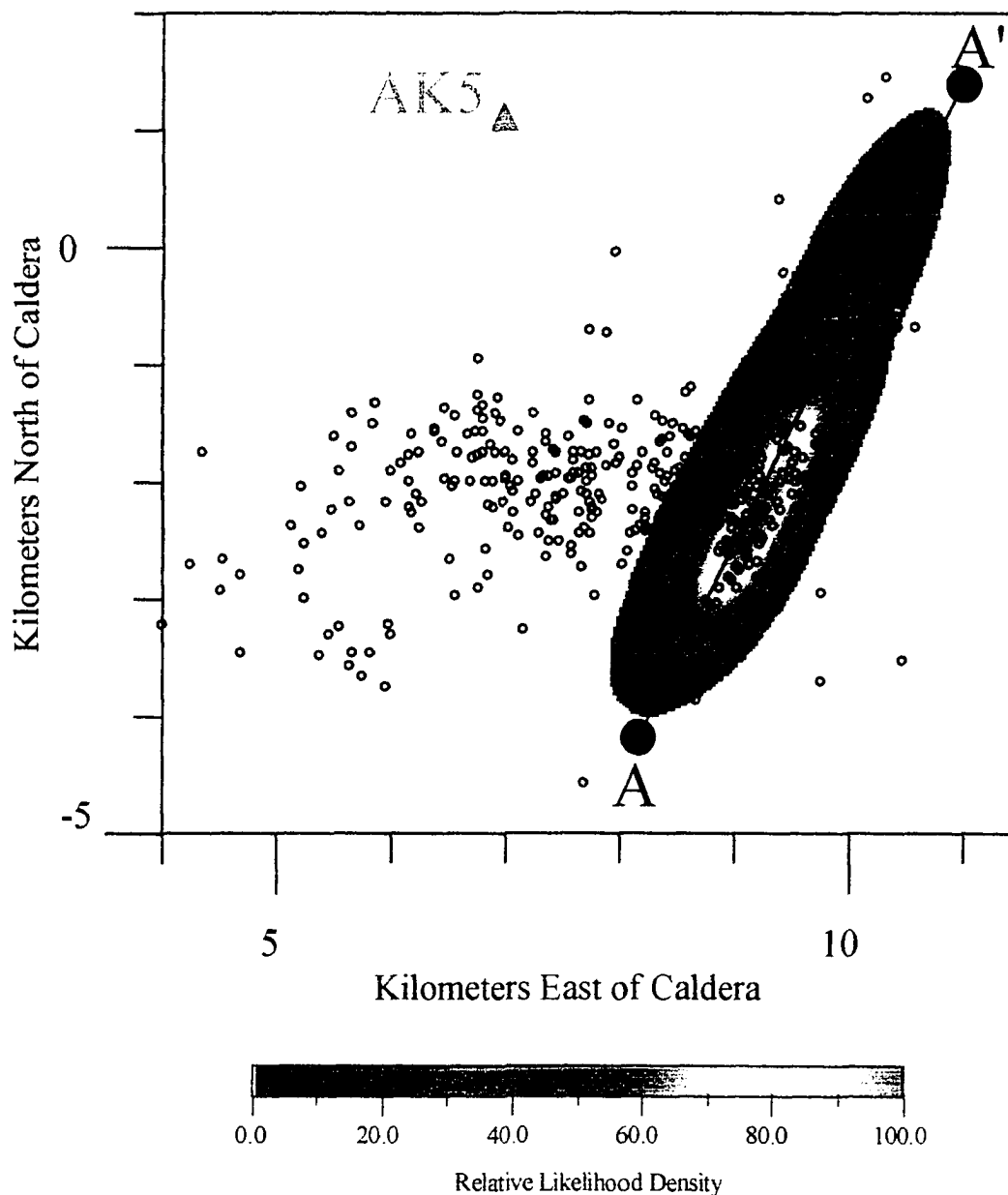


Figure 46: Tarantola-Valette L_2 norm epicentral density function and 432 maximum likelihood epicentral estimates in the detail area. The Tarantola-Valette epicentral density function was computed using modal arrival patterns and an estimated picking error of 0.1 second. Only epicenters for events with both S and P picks at all locations are included. Estimated picking errors of up to 0.5 seconds were allowed.

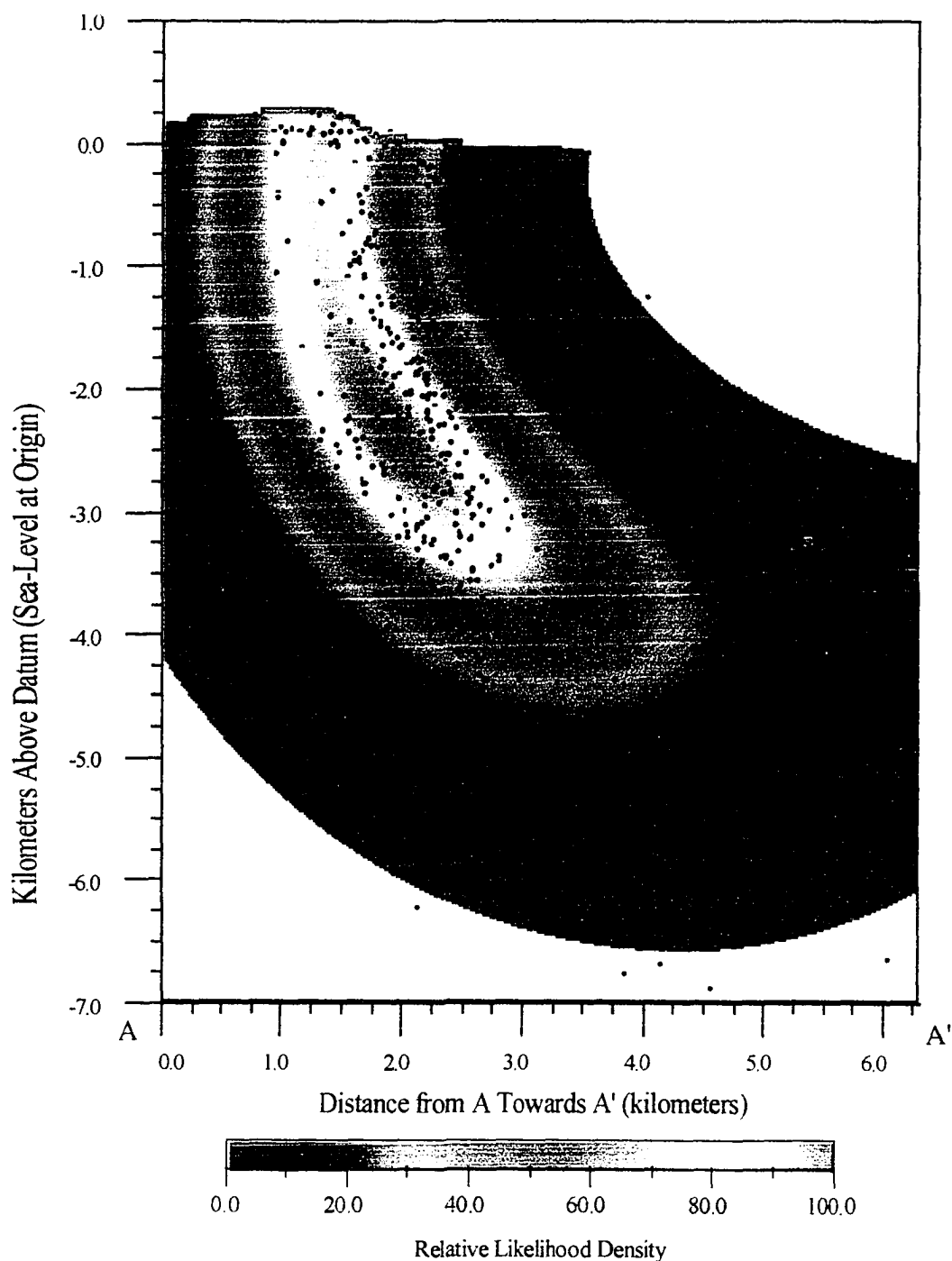


Figure 47: Tarantola-Valette L_2 norm hypocentral location density function in the vertical plane A-A' for dominant mode arrivals assuming a picking error standard deviation of 0.1 second at all locations. Estimated maximum likelihood hypocentral locations within 500 meters of the plane are projected normally into the plane, and are shown as points. Hypocenters were selected using the maximum likelihood location using the Tarantola-Valette L_2 norm.

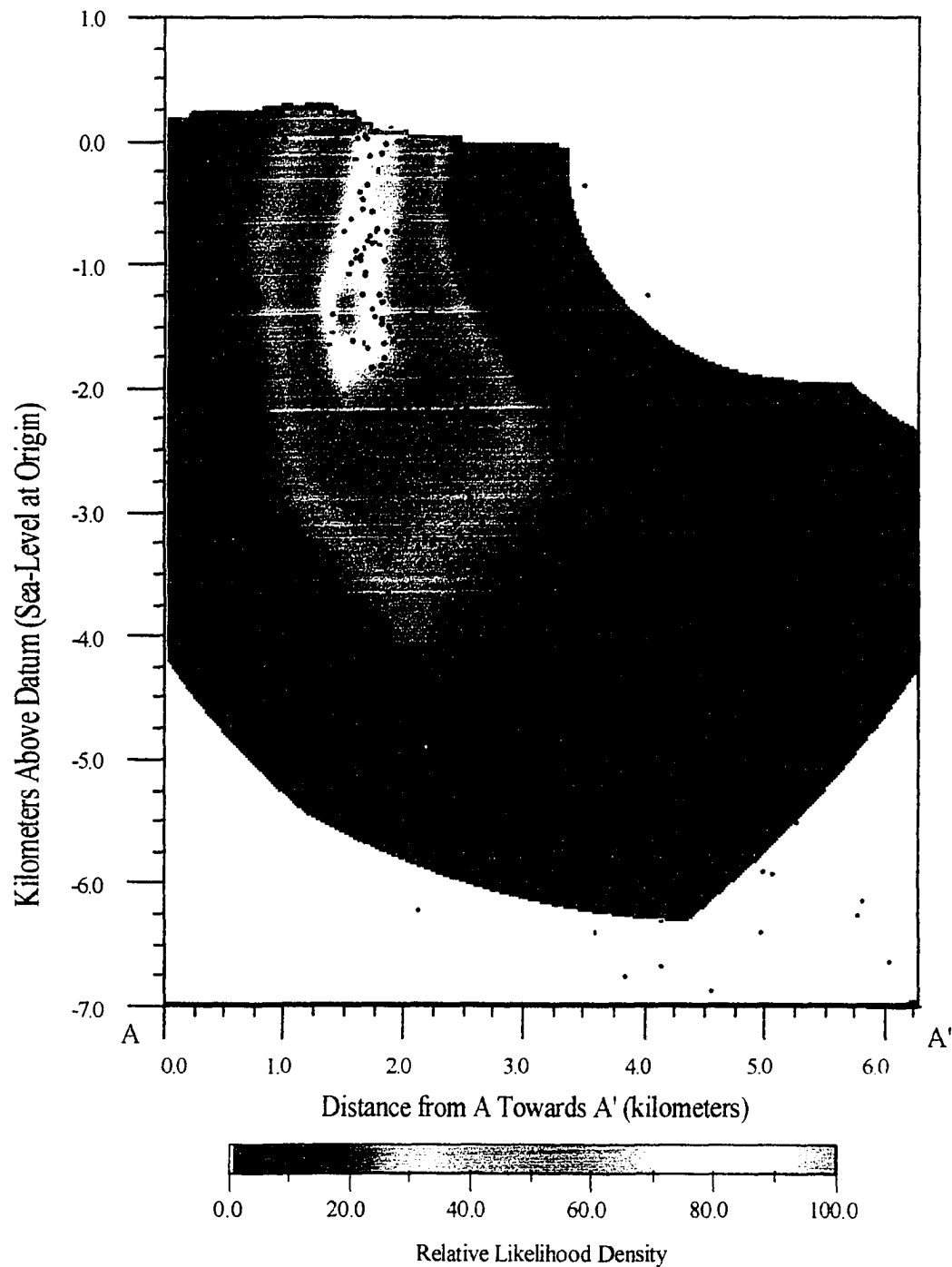


Figure 48: L_1 norm hypocentral location density function in the vertical plane A-A' for dominant mode arrivals assuming a picking error standard deviation of 0.1 second. Estimated maximum likelihood hypocentral locations within 500 meters of the plane are projected normally into the plane, and are shown as points. Hypocenters were selected using the maximum likelihood location using the Tarantola-Valette L_2 norm.

Epicentral Density Function and 432 Hypocenters in Detail Area

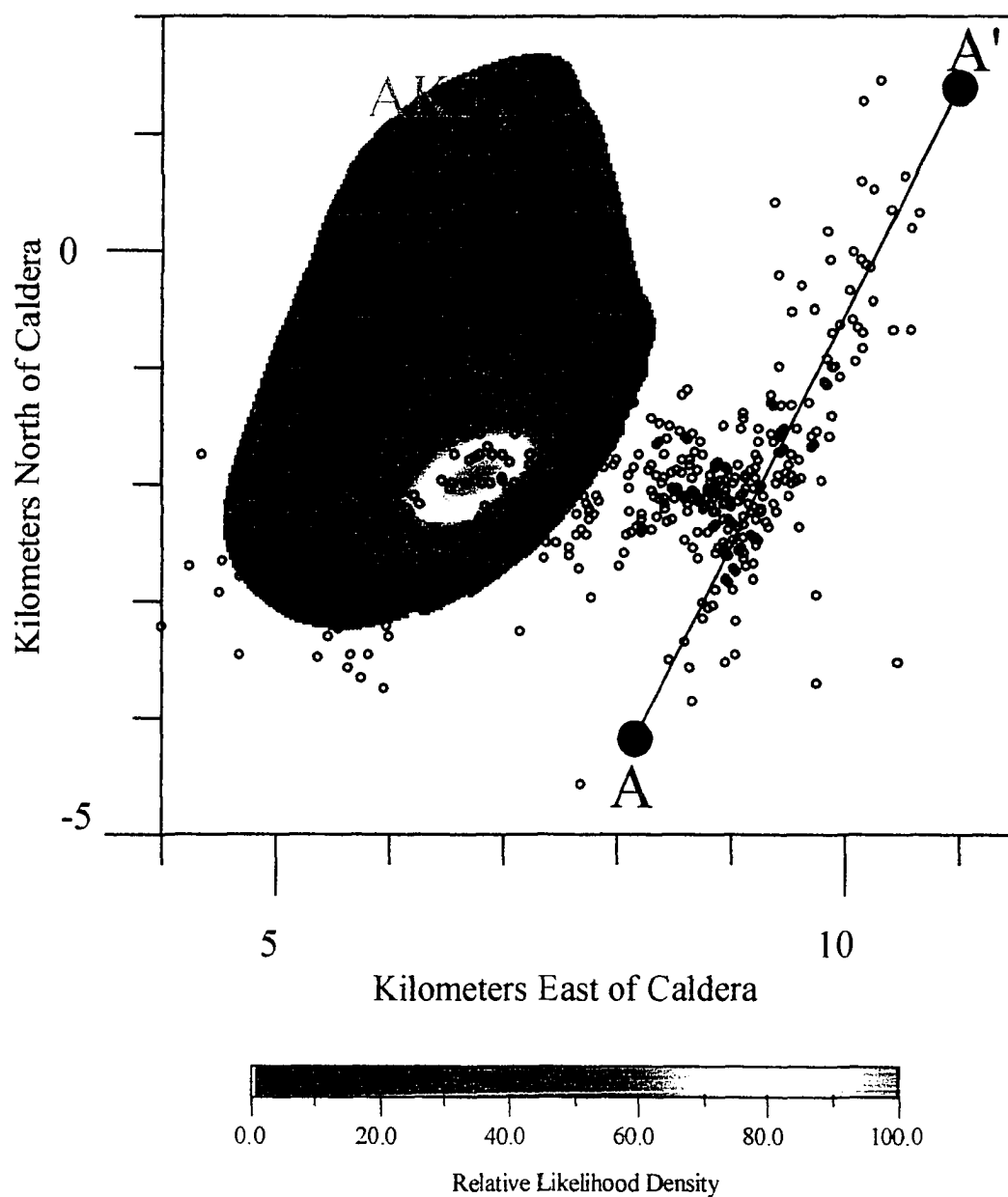


Figure 49: L2 norm epicentral density function for secondary mode arrivals and 432 maximum likelihood epicentral estimates in the detail area. The Tarantola-Valette epicentral density function was computed using secondary mode arrival patterns and an estimated picking error of 0.1 second. Only events with picking error estimates of less than 0.5 second for both *P* and *S* phases at all locations are shown.

Chapter 6

Seismic Tomography for Volcano Imaging: Design, Development, and Application to Redoubt, Mt. Spurr, and Augustine Volcanoes, Alaska

ABSTRACT

Tomographic methods used to image volcanic systems must address a number of challenges, including: 1) Robust and efficient raytracing through complex structural models; 2) Accurate location of hypocenters using realistic velocity models and sparse receiver arrays; 3) The ability to image the small, discrete features of geologic interest in volcanic systems. These challenges are large and it is tempting to proceed using a simplified implementation designed for less computationally challenging objectives. Many existing geotomography methods assume low contrast velocity models, accurate initial hypocenter location estimates, and ignore the issue of imaging small anomalies. We address each of these requirements, and develop an efficient tomographic imaging methodology that addresses the challenges of volcano tomography.

Raytracing is efficiently and robustly performed by first recovering approximate raypaths from pre-computed traveltimes fields. The raypaths are then refined to desired accuracy using modified pseudo-bending. Topography and large velocity contrasts are accommodated. Hypocenters are located using a coarse grid global search followed by nonlinear conjugate-gradient refinement with reset perturbations. Composite Distribution

Inversion (CDI) tomography is used in conjunction with smoothness constraints to permit estimation of large amplitude anomalies while still maintaining the stability of the inverse.

The implementation is demonstrated by deriving velocity models at Redoubt, Spurr, and Augustine volcanoes, Alaska. Our CDI images show a relatively small number of features and large velocity contrasts at depths of less than one kilometer below sea level. The deeper velocity structure is relatively homogeneous. In all three cases the RMS prediction error is reduced to a level approximately equal to phase picking accuracy. The velocity models are not unique and satisfy the data neither better nor worse than models derived using smoothness or damping constraints alone. One of two possibilities exist: 1) CDI suppresses anomalies at depth and exaggerates near-surface anomalies; or 2) Streaking associated with conventional constraints causes leakage of signal associated with strong near-surface anomalies into other portions of the model. Detailed absolute velocity information throughout the relatively shallow subsurface is required to identify the more correct interpretation.

INTRODUCTION

The objective of tomography is to image the interior of a body by inverting line-integrals passing through the body for media parameters (Gilbert, 1972; Hounsfield, 1973). Tomography has revolutionized biomedical imaging, and is occasionally applied in the field of geophysics (geotomography). The line-integrals used in seismic geotomography follow raypaths between seismic sources and receiving stations.

Traveltimes are used to estimate velocity structure, and amplitudes are used to estimate attenuation properties.

It is well known that biomedical tomography produces higher resolution images and more accurate body parameter estimates than seismic geotomography (Singh and Singh, 1991). This is primarily a consequence of practical acquisition geometry limitations intrinsic to most geotomographic imaging volumes. Biomedical targets permit experimental designs that produce well-posed inverse problems. Geotomograms, in contrast, are estimated from mixed-determined inverse problems.

Mixed-determined inverse problems have an infinite number of solutions that satisfy the observed data equally well. Stable point-solutions only exist after *a priori* information is imposed. Geotomographic methods most frequently add *a priori* information in the form of damping (Nolet, 1985; Lees and Shalev, 1992) or smoothness constraints (Sambridge, 1990). These constraints may be interpreted as model parameter distribution assumptions, and effectively stabilize, or 'regularize', the solution (Berryman, 1990). Regularization results in limited spatial resolution and highly attenuated anomalous amplitude estimates. Damping applied to crosshole tomography problems typically results in more than 60% attenuation of spatially small anomalies (Humphreys and Clayton, 1988; Clippard and Christensen, 1995). "Spatially small" means approximately the same size as the smallest anomaly that can be described by the model parameter basis functions. Natural source and 3D problems usually have higher levels of attenuation, and lower resolution. Lees and Crosson (1989) found that the best possible amplitude estimates in a relatively well resolved region of Mt. St. Helens might

approach 15% of the true value. This type of anomaly attenuation is caused strictly by the geometry of the problem and the information used to stabilize the solution. It is exacerbated by noise and poorly known source or station location parameters.

Volcanoes provide a unique tomographic imaging objective. High-resolution reflection and refraction methods are not viable volcano imaging tools because quasi-horizontal continuous interfaces do not approximate the expected structure of most volcanoes. Transmission tomography, on the other hand, does not depend on the existence of coherent reflecting or refracting horizons. Volcanoes often have abundant naturally occurring seismicity, and earthquake swarms are very common in volcanic settings (McNutt, 1996). These natural sources can be used to obtain denser raypath coverage and better suites of illumination angles than is possible in most geologic environments. A final unique aspect provided by volcanoes as a tomographic objective is provided by the topography of many volcanic cones. The existence of a small cone offers the potential to configure a group of receivers and artificial sources in a geometry similar to that used in biomedical tomography. Volcanic cones are one of the only geologic objectives that might feasibly permit such a configuration.

The opportunity to obtain detailed tomographic images of volcanoes also comes with a number of challenges. The velocity structure of many volcanic edifices is expected to be very heterogeneous. Raytracing methods must therefore perform robustly in the presence of arbitrary velocity field complexity, and must not fail in the vicinity of large velocity contrasts. Topographic relief at many volcanoes ranges from 1–4 km above the surrounding terrain. Raytracing methods employed for volcano tomography must

robustly accommodate such topographic relief. Tomography algorithms used for volcano imaging must also permit estimation of images consistent with the type of anomalies expected. Some features of critical interest, such as magmatic conduits, may be small relative to realizable spatial resolution. If such features are to be successfully imaged, the *a priori* information used to stabilize the inverse must permit the existence of spatially small, high-amplitude, discrete anomalies.

The objective of this paper is to: 1) Describe our approach to seismic volcano tomography; 2) Present velocity tomograms for, Redoubt, Spurr, and Augustine volcanoes, Alaska. The primary anomalous features are briefly discussed within the context of the corresponding volcanic system. Detailed volcanological interpretation of the models is beyond the scope of this paper.

METHODS

Seismic traveltimes tomography is performed by inverting a set of observed traveltimes for model parameters. The goal is to estimate a set of model parameters so that theoretical traveltimes are consistent with observations. Seismic tomographic inversion is intrinsically nonlinear because raypaths and traveltimes are nonlinear functions of the velocity field parameters. The nonlinear problem is usually solved as an iterative sequence of linearized steps from an initial solution towards progressively better solutions until convergence is reached. This approach is used because there are typically too many free parameters to feasibly permit a full nonlinear approach. We adopt this procedure, and iteratively solve the linear system

$$\begin{bmatrix} \mathbf{W} \mathbf{A}^{(k)} \\ \mathbf{S}^{(k)} \\ \mathbf{C}^{(k)} \end{bmatrix} \Delta \mathbf{m}^{(k)} = \begin{bmatrix} \mathbf{W}(\mathbf{t} - \mathbf{h}^{(k)}) \\ \mathbf{m}^{(0)} - \mathbf{m}^{(k)} \\ 0 \end{bmatrix} \quad (51)$$

where

k is the iteration number

\mathbf{t} is the vector of observed traveltimes

$\mathbf{h}^{(k)}$ is a vector of theoretical traveltimes at the beginning of iteration k

$\mathbf{A}^{(k)}$ is a matrix with elements $A_{ij}^{(k)} = \frac{\partial h_i^{(k)}}{\partial m_j^{(k)}}$

\mathbf{W} is a diagonal matrix with elements $w_{ii} = (\lambda_{i_p}^2 + \lambda_{i_m}^2)^{-\frac{1}{2}}$

$\mathbf{m}^{(0)}$ is the *a priori* estimate of parameter values, i.e. the reference model

$\lambda_{i_p}^2$ is the variance of the i^{th} observed arrival time

$\lambda_{i_m}^2$ is the variance of the i^{th} theoretical modeling error

$\mathbf{S}^{(k)}$ provides a set of smoothness or flatness constraints for $\mathbf{m}^{(k)}$

$\mathbf{C}^{(k)}$ is a diagonal matrix with elements $c_{ii}^{(k)} = \frac{1}{\sigma_i^{(k)}}$, off-diagonals = 0

$\sigma_i^{(k)}$ is the *a priori* expected standard deviation of the i^{th} parameter at iteration k

and $\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \Delta \mathbf{m}^{(k)}$ is the $(k+1)^{\text{th}}$ estimate of the model parameter vector.

Equation (51) is iteratively solved until convergence criteria are satisfied. The vector

$\mathbf{m}^{(0)} - \mathbf{m}^{(k)}$ can be set to the null vector if the damping factor is used only as a

stabilization term. This results in an inverse that stabilizes perturbations relative to the

current model rather than measuring the solution distance from the initial reference

model.

There are five components that strongly influence the solution produced by the above procedure:

- 1) The availability of sufficient data;
- 2) Accuracy of theoretical traveltimes and their partial derivatives;
- 3) The type of *a priori* information added to the problem;
- 4) Model parameterization, free parameter selection, and solution strategy;
- 5) Robustness of the large, sparse linear-systems solver used to solve equation (51).

Data Requirements and Selection

Reliable tomograms can only be calculated if a large number of raypaths sample the model parameters upon which material property calculations are based, throughout the imaging area of interest. Many illumination angles should be present in the primary imaging area. Source and arrival times should be well known, as should source and receiver locations (Chapter 4, this thesis).

When earthquake sources are used, source times and locations are estimated using a velocity model. This results in a coupled inverse problem of hypocenter and velocity model parameter estimation. If a large number of observations are available, the optimal source location estimate may be relatively insensitive to minor velocity model perturbations. This permits linearization of the hypocentral estimation problem in the neighborhood of an initial estimate. Optimal velocity and hypocenter parameters can then be simultaneously estimated (Pavlis and Booker, 1980; Abers, 1994).

The accuracy and robustness of an earthquake source location estimate is primarily a function of the number and quality of picks available, the aperture of the array with respect to the location of the source, and the suitability of the velocity model used. Accurate traveltimes calculation is assumed. A measure referred to as the *gap* is often used as an indicator of the array aperture with respect to the source location (Lahr, 1980). Gap is defined as the maximum difference between adjacent azimuths from the estimated source to the stations used to estimate the location.

We select data based on the number and quality of available picks, and the gap. A minimum depth relative to the distance to the closest station is sometimes also used as a criterion for data selection (Villaseñor et al., 1997). This helps prevent topographic interference with raypaths, particularly *S* phases which may be strongly coupled with the free surface. We do not use this criterion, primarily because there are large number of events within the volcanic cone at the locations investigated here. Sources within the volcanic cone can provide detailed information about velocities in the cone itself.

The volcano tomograms presented here use earthquakes as sources, however only a marginal number of phase observations from 6 – 9 stations is the rule (more stations are used for a three-month period of time at Redoubt). Small arrays tend to produce highly nonlinear hypocentral location density functions, and linearized estimation of optimal hypocentral parameters is often inadequate (Moser et al., 1992; Zhou, 1992; Chapter 5, this thesis). As a consequence, we approach the problem by iteratively estimating optimal hypocentral parameters and velocity model perturbations until convergence of both the velocity model and the hypocenter parameter estimates is reached. A fully

nonlinear method is used to revise hypocenter location estimates at each iteration (Chapter 5, this thesis). For data volumes that usually include 8 or more phases per event, we estimate source locations as the L_2 norm maximum likelihood location. For data volumes that tend to have fewer than 8 phases per event, we use the L_1 norm, which produces more robust hypocentral estimates for small or noisy data sets (Thurber, 1985; Moser et al., 1992). We have attempted to mix the norms on an event-by-event basis within a data volume, but found the difference in optimal location estimates as a function of the norm alone was too large to permit use of both location estimates in the same tomographic inversion.

A final data requirement is that the sources should be uniformly distributed spatially. Good spatial distribution of sources ensures many illumination angles throughout the volume, and helps to maintain near equality of the columns of $\mathbf{W} \mathbf{A}^{(k)}$ in equation (51). Linear systems solvers perform best when the elements in each non-null column of $\mathbf{A}^{(k)}$ have equal lengths under the L_2 norm (Paige and Saunders, 1982). Direct column scaling after constraints have been added is not possible because such scaling alters the physical interpretation of the constraints. It is therefore expedient to condition the input data so that the columns of $\mathbf{W} \mathbf{A}^{(k)}$ are as nearly uniform in magnitude as possible. This can be accomplished either by discarding closely spaced hypocenters (Benz et al., 1996), or by down-weighting events that are closely spaced based on a measure of spatial correlation. We use the latter approach, and pre-multiply the matrix \mathbf{W} by a diagonal correlation scaling matrix \mathbf{V} . The diagonal elements of \mathbf{V} with diagonal elements v_{ii} given by

$v_{ii} = \left[\sum_j^M \text{cor}(t_i, t_j) \right]^{-\frac{1}{2}}$. The function $\text{cor}(t_i, t_j)$ is a measure of the correlation between

the i^{th} and j^{th} observations, and should range in value from 0 to 1. The sum of squared error contributions from a group of perfectly correlated observations (i.e. identical source and station locations) would make the same contribution to the overall prediction error as a single independent observation, but the noise within the observations would be

averaged within the group. One possible form for $\text{cor}(t_i, t_j)$ using just the source and

receiver locations is to set $\text{cor}(t_i, t_j) = \exp \left[-\frac{1}{2} \frac{\|\mathbf{x}_i^{(\text{source})} - \mathbf{x}_j^{(\text{source})}\|^2 + \|\mathbf{x}_i^{(\text{station})} - \mathbf{x}_j^{(\text{station})}\|^2}{\Delta^2} \right]$,

where Δ is some *a priori* correlation distance. Another possibility is to use a measure of

correlation between the rows of \mathbf{A} , so that $v_{ii} = \left[\sum_j^M \text{correlation}(\mathbf{A}_i^T, \mathbf{A}_j^T) \right]^{-\frac{1}{2}}$. Our

informal experiments that there is little practical difference in the final results generated as a function of the form of down-weighting selected. Data discarding is also an acceptable option, and reduces the size of the linear system.

Traveltime Calculation and Velocity Model Basis Function Selection

Many methods are available for calculating theoretical seismic traveltimes (Julian and Gubbins, 1977; Cervený and Psencik, 1983; Keller and Perozzi, 1983; Langan et al., 1985; Cervený, 1985; Um and Thurber, 1987; Hanyga, 1988; Prothero et al., 1988; Virieux et al., 1988; Vidale, 1988, 1990; Sambridge and Kennett, 1990; Moser, 1991;

Podvin and Lecomte, 1991; Schneider et al., 1992; Klimes and Kvasnicka, 1994; Grechka and McMechan, 1996). We use ray theoretic traveltimes with ray geometry controlled by the eikonal equation, $\frac{d}{ds} \left(\frac{1}{V} \frac{d\mathbf{x}}{ds} \right) = \nabla \left(\frac{1}{V} \right)$ (Aki and Richards, 1980). Ray theory is based on the assumption of infinite frequency wave propagation (Wielandt, 1987) and assumes that the arrival actually observed is associated with the minimum time raypath. It is well known that this approximation may not always be valid (Geoltrain and Brac, 1993) and that imaging based on first arrivals is limited, particularly when traveltimes are based on gridded eikonal solvers (Gray and May, 1994). Imaging accuracy can be improved in some cases by using maximum energy wavepaths (Nichols, 1996). Maximum energy wavepath calculation is much more computationally demanding than minimum time raytracing, and we adopt the minimum time approach primarily for that reason.

Within the context of minimum time raytracing, many computational options still exist. We use an approach based on initial approximate raypath recovery from an approximate traveltimes field, followed by raypath refinement to required accuracy (Chapter 4, this thesis). The method is sufficiently accurate, and robustly handles velocity models of arbitrary complexity. Raytracing failure due to the presence of local caustics (Abers, 1994) does not occur, and head waves can be correctly predicted as first arrivals. Topographic interference with raypaths is accommodated within the limitations of the minimum time raypath assumption. This may be the greatest weakness of the method since P arrivals observed in areas of rugged topography are likely to be

refractions from deeper horizons, rather than arrivals that traveled through a minimum time, but high energy loss, path. S phases may be even more strongly influenced by topography than P waves because shear waves can couple strongly with the free surface and propagate at a group velocity that is quite different from the body wave V_s in the area.

The accuracy of traveltime partial derivatives with respect to model parameters is largely a function of the consistency of the velocity model parameterization with the method used to calculate traveltimes. Constant velocity cube-shaped basis functions (square in 2D) are commonly used to describe media properties and seem to produce adequate results (Vasco et al., 1996). The primary inadequacy associated with cube-shaped basis functions is that the velocity gradients estimated are necessarily inconsistent with the underlying model because the actual gradients are either zero or infinite. Both ray-bending and finite difference traveltime field extrapolation use velocity gradients to calculate traveltimes. It is not desirable to trace rays by applying Snell's Law at the actual hard boundaries of the basis functions because the predicted ray trajectories are extremely sensitive to the (artificial) orientation of the block boundaries

Another more subtle inconsistency can occur when non-overlapping basis functions are used in conjunction with optimal hypocenter location estimates using traveltimes derived directly from the traveltime fields. If the hypocenters are optimal with respect to a set of traveltime fields calculated using gradient approximations, but line integrals along the raypaths are used to estimate partial derivatives with respect to model parameters, residuals associated with the method inconsistency arise. These residuals

cause some component of the final velocity model estimate to be an image of the traveltimes calculation inconsistency. The error magnitude can range up to a few percent, which is not necessarily negligible considering the very limited range of anomalies that can be successfully imaged tomographically.

For the reasons outlined above, we do not use constant velocity cube-shaped basis functions. All velocity models presented in this paper use basis functions with triangular profiles and 50% overlap. Velocities are calculated using tri-linear interpolation. Velocity gradient and raypath segment traveltimes calculations are consistent with the underlying velocity model. Traveltimes fields are only used to predict approximate raypaths. The actual traveltimes used to estimate optimal hypocenter locations are based on refined raypaths identical to those used when calculating partial derivatives.

Another advantage of overlapping triangular basis functions is that fewer model parameters are required to describe a smooth model. This is because a smooth model can be adequately described using a larger grid node spacing, Δx , than is required by constant velocity basis functions. Any parameterization that reduces the number of parameters required while still providing an adequate physical description is desirable because the resulting problem is usually smaller and better conditioned. The total number of model parameters, which is proportional to Δx^{-3} for regular 3D grids, can often be reduced substantially without reducing model fidelity.

***A Priori* Information**

The type and accuracy of the *a priori* information used largely controls the nature of all seismic tomograms. The single most important piece of *a priori* information is the assumed reference model. As indicated earlier, all parameter estimates are actually perturbations away from the *a priori* reference model. Almost none of the parameter perturbations are uniquely determined. It is therefore absolutely paramount to establish the best possible reference model prior to estimating a detailed tomographic image. We use a minimal 1D inverse as the reference model (Kissling et al., 1994). The minimal 1D inverse itself can have a non-unique solution, however. This is particularly true if deep events have been mislocated using an incorrect model and are then included in the initial 1D estimation procedure. Shallow velocity estimates can then be strongly biased towards a 1D solution that is compatible with the initial deep mislocations.

We estimate *a priori* reference models by first inverting for shallow, 1D velocity structure using only the most shallow earthquakes, together with any shots that might be available. Station corrections are simultaneously estimated. The depth of the estimated 1D structure is then extended by sequentially inverting an ever deepening population of events. At each step, station corrections are revised. One-dimensional smoothness constraints are used, when required, to obtain both smooth and well-determined solutions. This approach reduces the bias of shallow parameters caused by use of a poorly known initial model.

The type of *a priori* information used to stabilize, or ‘regularize’, the inverse also plays a critical role in determining the nature of tomograms actually estimated. Simple

damping is the most easily implemented and most computationally efficient option. Uniform damping requires no extra data storage, and most 'canned' methods of solving linear systems allow simple uniform damping as an option. Damping amounts to making an assumption of model parameter perturbation variance. The solution is a weighted measure of L_2 model perturbation length and minimum prediction error (Menke, 1984). Damping strongly attenuates large amplitude anomalies and does not include facilitate specification of model parameter covariance (Clippard and Christensen, 1995).

Smoothness or flatness constraints are often used to stabilize tomographic inverses (Sambridge, 1990; Lees and Shalev, 1992). These constraints amount to an assumption of covariance between adjacent model parameters. This is desirable, since the scale of tomographic basis functions is usually selected to be smaller than the actual potential resolution in order to accommodate smoothly varying velocity models.

Composite distribution inversion (CDI) constraints were introduced to permit detection of large amplitude, but spatially small, anomalies in a well known background (Clippard and Christensen, 1995). CDI constraints are implemented by iteratively modifying the *a priori* expected variance of each parameter using its probability of belonging either to a background or an anomalous model population. The probability of belonging to background or anomalous populations is established using a parameter estimate from a previous iteration. The full advantage of CDI constraints is only realizable when the reference model is well known. Other methods of mitigating the deleterious effects of regularization have been proposed, and positive results have been shown (Carrion 1991; Hyndman et al., 1994; Vasco et al., 1996; Hyndman and Harris,

1996; Carrion et al., 1997). We believe that CDI offers performance comparable to the other methods with less specific *a priori* information.

We use a combination of smoothness and CDI constraints. The velocity models we estimate here have poorly known reference models, and parameters selected for full CDI anomaly detection do not produce reliable results. We therefore choose CDI parameters that allow higher amplitude anomaly estimates than damping might, while still permitting a large fraction of parameters to be anomalous. Smoothness constraints are used to maintain spatial anomaly coherence. The magnitude of the smoothness and CDI constraints selected is interpretive, and must be chosen to produce perturbation estimates that are within an acceptable *a priori* range, while effectively reducing prediction error.

Parameterization of the Primary Free Parameters

Traveltime tomography can be used to estimate *P* and *S* wave velocities simultaneously or independently. Station corrections may be simultaneously estimated with the velocities, held fixed, or updated using a measure of central tendency after estimating the other parameters. Hypocenter location estimates may be revised between velocity estimate iterations, or perturbations to the location can be derived simultaneously with the velocity parameter estimates using a linearized approximation of the optimal location perturbation. The velocity model can be parameterized directly in terms of velocities or slowness, Lamé constants, or coefficients of basis functions from which the desired physical parameters may be computed. We have experimented at length with various parameterizations. A complete review of the results is beyond the scope of this

paper. We only describe the approach used to generate the results shown here, together with the primary rationale for choices made.

We choose to parameterize our models directly in terms of velocities at specific spatial locations. A seeming disadvantage is that the perturbations to be estimated are made using a sensitivity matrix composed of derivatives truncated after the first term of a Taylor series expansion. Parameterization in terms of slowness does not require truncation of the partial derivatives, and might be considered preferable. In practice, there is little advantage to using slowness during the inversion step because the raypaths from which the partial derivatives are derived are nonlinear functions of the slowness coefficients. The primary disadvantage to using a slowness parameterization is that the gradients used to estimate raypaths are necessarily first order truncated approximations. We find that the negative influence of linearizing the velocity gradient approximation during raytracing is a more significant concern than linearizing the traveltime partial derivatives during construction of the sensitivity matrix. We therefore parameterize models in terms of velocity.

When sufficient reliable phase arrival information exists, both P and S wave velocities can be estimated. When this is possible, we choose to simultaneously estimate parameters controlling both V_p and V_s simultaneously. The actual parameters we estimate are V_p and $\frac{V_p}{V_s}$. The reason for this approach is that $\frac{V_p}{V_s}$ is often a roughly constant and fairly well known quantity, even when absolute velocities are poorly known. This correlation should be incorporated into the solution as some form of *a priori*

information. It is possible to estimate either V_p or V_s first, and then specify constraints while estimating the other to produce results in the neighborhood of the prior $\frac{V_p}{V_s}$ value. This approach, however, provides incomplete coupling of available information because the P raypaths are different from the S raypaths. The resulting difference in the structure of the sensitivity matrices is more prone to generate $\frac{V_p}{V_s}$ anomalies that are caused by difference in ray coverage alone than if both parameters are simultaneously estimated. It is also possible to specify the prior relationship of V_p and V_s such that both V_p and V_s are directly estimated, although appropriate scaling of the constraint system is more difficult than constraints applied to $\frac{V_p}{V_s}$ directly. Estimation of Lamé constants instead of velocity parameters, either with or without $\frac{V_p}{V_s}$ assumptions, offers no intrinsic advantage to the approach used here.

Before estimating 3D solutions, a minimal 1D inverse should be derived for use as both an initial and reference model (Kissling et al., 1994). We derive the minimal 1D inverse using a procedure identical to that used for full 3D imaging. We make no computational distinction between 1D, 2D, and 3D tomography, and the core of our implementation makes no specific reference to any velocity model representation. Any velocity model capable of providing velocities, gradients, and partial derivatives with respect to parameter indices can be used as a model to be optimized with respect to a group of phase observations. The velocity models themselves are responsible for both

constructing the constraint systems used during inversion, as well as providing the parameter revision services required after a perturbation estimate is made. The 1D models used here all specify either V_p and $\frac{V_p}{V_s}$ at a set of tie points at various depths, or V_p and a constant $\frac{V_p}{V_s}$ for the entire model. Values of the parameters are linearly interpolated between tie points.

Use and Estimation of Station Corrections

Station corrections account for consistent differences between observed traveltimes and theoretical traveltimes predicted by a model for each station and phase used in the tomographic inversion. As such, the appropriate station correction is dependent upon the ability of the model to describe localized velocity anomalies in the vicinity of the station as part of the model, and the resolution scale of interest. For 1D models, we derive station corrections simultaneously with the velocity parameters. For 3D models, station corrections within the imaging area are set to zero, which forces anomalous traveltimes to predict anomalous velocities near the station. We feel that this is justified given our actual interest in near surface velocities and the small scale of the models used here. Corrections for stations outside the model space are derived simultaneously with the model parameters. This maps all consistent velocity anomalies outside the volume imaged into a lumped station correction.

Note that some station corrections might be appropriate due to reported station location inaccuracy or inconsistency. Some station locations are based on map coordinates, while others are based on GPS coordinates. These issues are not addressed in this paper. We did not consider the magnitude of the differences associated with the datum inconsistency sufficiently large to justify an effort to estimate the appropriate corrections.

Estimation of Earthquake Source Locations

Earthquake locations are derived using a full nonlinear optimization procedure between velocity model estimation iterations (Chapter 5, this thesis). A mixed system of source location perturbations and velocity parameters is not constructed, and they are not simultaneously estimated. We use this approach for three reasons.

First, we find a large amount of nonlinearity in the hypocentral location cost function for the small arrays typically used for volcano monitoring (Chapter 5, this thesis). Optimization of the hypocentral location estimate after a velocity perturbation can require several downhill line searches over a substantial distance from the original location. On the other hand, the large number of parameters in the velocity model allows effective residual reduction with a single perturbation estimated using linearization. If the linearized estimates of the perturbed hypocentral approximations were accepted, residuals associated with the non-optimal location with respect to the current velocity model would map into unnecessary velocity model perturbations.

The second reason to avoid linearizing the location problem is provided by our observation that analytic derivatives of the hypocentral cost function in heterogeneous models can be quite different from the numerical differences actually observed. The region over which the analytic first derivative correctly predicts the change in cost function perturbation can be small relative to the distance required to reach a truly optimal point. The largest differences occur in regions where raypaths change substantially with a small change in source location. This happens frequently in strongly heterogeneous 3D models.

The third reason to decouple source location estimation from velocity parameter estimation is that a modeling error term should be used during location to account for velocity model inaccuracy. The term helps to provide robust location estimates in overly simplified models. This term should *not* be included during velocity parameter estimation, because what is ‘model-noise’ to the location problem is in fact the desired input signal to the velocity parameter inverse. The actual cost function we use for location is of the form

$$E(\mathbf{x}) = \mathbf{r}^T (\mathbf{C}_d + \mathbf{C}_m)^{-1} \mathbf{r} \quad (52)$$

where \mathbf{r} is a vector of weighted generalized prediction errors, \mathbf{C}_d is the data covariance matrix, and \mathbf{C}_m is the modeling error covariance matrix. The matrix \mathbf{C}_d is mostly diagonal, except for observations associated with the same source and station location. This occurs in the case of three component or colocated stations. Matrix \mathbf{C}_m accounts for correlated modeling errors, which are primarily associated with velocity model inadequacy. We

choose the elements of \mathbf{C}_m to be given by $c_{ij} = \sigma_i \sigma_j \exp\left(-\frac{1}{2} \frac{D_{ij}}{\Delta^2}\right)$, where σ_i is the modeling error of the i^{th} ray, D_{ij} is the distance between stations observing the i^{th} and j^{th} rays, and Δ is a correlation length (Tarantola and Valette, 1982). During initial location with a 1D model, Δ is chosen to be relatively large. When estimating source locations using detailed three-dimensional models Δ is chosen to be small (on the order of the grid spacing). A modeling error term may be included during estimation of velocity parameters (51), but the appropriate correlation distance is quite different from that used during source location estimation.

Coordinate System

The models presented in this paper are sufficiently small that a flat earth model is adequate. Nevertheless, we designed our tomography algorithm for use at any scale. All calculations are performed in a rotated and translated Cartesian coordinate system to avoid scale distortion associated with geographic coordinate projections. Given a latitude and longitude near the center of the imaging volume, an origin is selected at that latitude and longitude at sea level. A local coordinate system is defined at this point using some specified ellipsoid, the z-axis is pointed upward, the y-axis is pointed toward geographic north at the origin, and the x-axis is pointed east. The x-y plane is tangent to the ellipsoid at the origin, and sea level always has a z-coordinate less than or equal to zero. Depth dependent velocities are derived by using a depth given by the actual distance from the

desired point to the ellipsoid surface along the surface normal passing through the desired location.

All models presented here are based on the GRS1980 ellipsoid (Snyder, 1982). Distances indicated on figures are presented in the undistorted local model space, rather than as projected geographic coordinates.

Solving the Linear System

A number of large, sparse, linear systems solvers are available. Singular value decomposition (SVD) is the most robust, generates condition numbers, identifies the dimensionality of the null space, and facilitates computation of resolution matrices (Lanczos, 1961; Golub and Reinsch, 1970; Lines and Treitel, 1984; Lines and LaFehr, 1989). Unfortunately, the size of most 3D tomography problems is too large to permit effective use of SVD. Row-action methods, such as SIRT (Gilbert, 1972; Ivansson, 1986) and ART (Kaczmarz, 1937; Tanabe, 1971; Dines and Lytle, 1979), are efficient, but generate very little diagnostic information and are somewhat difficult to control. Linear conjugate gradient methods (Hestenes and Stiefel, 1952; Paige and Saunders, 1982; Scales, 1987) are also very efficient, provided a method of sparse matrix storage that permits rapid recovery of both row and column vectors. Of the linear conjugate gradient methods, methods using the Polak-Ribiere modification (Polak, 1971; Gill et al., 1981), such as LSQR (Paige and Saunders, 1982), are found to estimate solutions similar to those obtained via SVD, with computation times similar to ART or SIRT (Nolet, 1985; Benz et al., 1996). We have experimented with the above methods using an abstracted

C++ implementation that permits full interchange of linear system solvers with no other implementation differences. We find linear conjugate gradient is easier to reliably use than ART or SIRT, and that it converges more quickly toward the SVD solution. Performance can be enhanced if the sensitivity matrix and its transpose are both stored using sparse vectors, which allows very efficient row and column vector multiplication. The results presented here are generated using our own C++ (Stroustrup and Ellis, 1994) implementation of the LSQR algorithm. Both the sensitivity matrix and its transpose are stored when sufficient computer memory is available.

DATA

General

Tomograms are computed at Redoubt Volcano, Mt. Spurr, and Mt. St. Augustine. All are active holocene volcanoes in the Cook Inlet region of south-central Alaska. Mt. St. Augustine last erupted in 1986, Redoubt in 1989-90, and Spurr in 1992. The volcanism is related to the northward subduction of the Pacific plate beneath Alaska.

Naturally occurring earthquakes generated most of the raypaths used to produce the velocity models presented here. Chemical explosion sources (referred to throughout this paper as 'shots') supplement the earthquake data at both Redoubt and Augustine.

Most of the waveform data were transmitted using analog telemetry with a dynamic range of approximately eight bits. Data were generally recorded digitally at a sampling interval of 0.01 second. Phase arrivals were identified by analysts representing the USGS, the University of Alaska Fairbanks Geophysical Institute Seismology Laboratory,

the Alaska Volcano Observatory, or some combination thereof. Exceptions to these generalizations are noted below.

Arrival onset identification accuracy (picking accuracy) for permanent array data is estimated to be no better than 0.06 seconds for *P* phases, and 0.1 second for most *S* phases. The actual estimated picking accuracy for all data was lost during data storage. Only a *weight code*, which indicates an upper limit of the original estimated picking error, was retained. Upper limit boundaries for weight codes (0,1,2, and 3) are (0.1, 0.5, 1.0, and 2.0) seconds, respectively, for most of the data. More accurate picks are available for 218 events at Redoubt. Most raypaths used in this study have total traveltimes of less than 2.0 seconds. The coarsely spaced weight codes relative to the scale of the imaging objective and raypath traveltimes does not permit accurate estimates of source location uncertainty.

Data at all locations are selected based on a number of criteria, one of which is the estimated picking accuracy. *P* phases are only considered to be *well picked* if the estimated picking accuracy is less than or equal to 0.1 second. *S* phases are considered to be well picked if the estimated picking accuracy is less than or equal to 0.5 second. Worse quality picks are also considered, with appropriate weighting, for qualifying events. Events are selected based on the number and quality of available picks, and gap. Events are not rejected because of large RMS error. Selection of only events with small RMS error effectively pre-selects data that is compatible with the existing velocity model, leaving little residual to indicate the existence of perturbations.

Redoubt Data

The United States Geological Survey (USGS) detonated four shots at Redoubt Volcano on July 10-11 1991 (Benz et al., 1996). A temporary array of 20 three-component stations recorded both the Redoubt shots and arrivals from local volcano-tectonic earthquakes during a three-week deployment period. The four shots, plus picked phases from 218 volcano-tectonic (VT) events recorded with the temporary array are considered for use in this study. The best picks from the temporary Redoubt array may be as small as 0.02 seconds. Upper limit boundaries for weight codes (0,1,2, and 3) are (0.02, 0.05, 0.1, and 0.025) seconds, respectively.

Permanent array data recorded during the period January 31, 1990 and August 12, 1997 are also considered as potential sources. A combined total of 835 earthquakes, plus the four shots are used to estimate both P and S wave velocity structure at Redoubt Volcano. All events used in this study satisfy following criteria:

- 7 or more picked phases;
- 5 or more well picked P phases;
- 6 or more total well picked phases;
- A maximum gap of 270 degrees.

The total data volume consists of 5493 P phases and 5809 S phases. Many of the S phases were observed on more than one component of the same station and were appropriately weighted during both source location estimation and tomographic

inversion. Both the *P* and *S* phases are used for source location and velocity parameter estimation. Station and shot locations are shown in Figure 51.

Spurr Data

Earthquake sources alone are used to generate the velocity model at Mt. Spurr. A total of 556 events recorded between January 23, 1990 and August 18, 1997 are used. All events used analog telemetry and digital recording. The events used in this study satisfy following criteria:

- 7 or more picked phases;
- 5 or more well picked *P* phases;
- 6 or more total well picked phases;
- A maximum gap of 240 degrees.

The total data volume consists of 4284 *P* phases and 3302 *S* phases. The *S* phases are only used to constrain source location estimates. Efforts to produce a detailed 3D *S* velocity model suggest that there are insufficient high quality *S* picks to permit reliable estimation of detailed 3D *S* velocity structure. The *S* velocity model used for earthquake location was the *P* model divided by a constant $V_p/V_s = 1.8$. Station and shot locations are shown in Figure 63.

Augustine Data

In addition to earthquake sources, shot data were detonated at Augustine on July 1, 1977 (Pearson, 1977), and August 16-17, 1995. During the 1977 experiment, five

permanent stations and twelve temporary stations were used to record shots detonated in nine unique locations (Figure 70). Analog recording was used, and picking accuracy is estimated to be 0.1 second. Eleven shots were detonated in 1995. The number of stations at which arrivals could be identified varied from 7-10. These shot data were recorded using analog telemetry followed by digital sampling at a rate of 0.001 seconds.

The original goal of the 1995 experiment was to deploy a dense, temporary array of high-frequency seismometers and a digital data acquisition system. This system would have produced over 7000 raypaths and picks as accurate as 0.002 seconds. Unfortunately the system was not deployed due to weather conditions during the period of instrument availability. Actual picks for the 1995 shot data are estimated to be no more accurate than 0.06 seconds.

Augustine earthquake data recorded during the time period of May 18, 1993 to August 9, 1997 are also used to provide raypath information. The data were transmitted using analog telemetry and recorded digitally. All events used in this study satisfy following criteria:

- 5 or more picked phases;
- 3 or more well picked *P* phases;
- 4 or more total well picked phases;
- A maximum gap of 270 degrees.

A total of 102 earthquakes provide raypaths used during inversion. The total data volume consists of 785 *P* phases and 389 *S* phases. The *S* phases are only used to constrain source location estimates. We do not believe there are sufficient high quality *S* picks to

permit estimation of detailed 3D S velocity structure, and none is presented. The S velocity model used for earthquake location is the P model divided by a constant $V_p/V_s = 1.85$.

RESULTS

The results presented here are not unique and are based on interpretive selection of constraint parameters that permit effective reduction of prediction error without obviously spurious parameter estimates. The parameters used are a function of the units used during inversion. The actual velocity parameters we estimate are numerically specified in terms of km s^{-1} . Spatial coordinates are specified in kilometers, and time units are specified in seconds. V_p/V_s parameters are dimensionless. The constraints reported here are appropriate for these units. Duplication of the results shown here would require constraint scaling if a different set of units were used.

Redoubt

The Redoubt shot experiment provides absolute velocity information for the portion of the volcanic edifice rising above sea level. This shot information, combined with earthquake data recorded by the temporary array, permits a relatively unique determination of absolute velocity in the shallow subsurface. This model, in turn, permits better source location estimates for deeper earthquakes, and better determination of velocities throughout the model space. Our estimate of a one-dimensional P-wave

velocity structure is similar to the previous one-dimensional layered structure used by the Alaska Volcano Observatory (Figure 53). The one-dimensional S-wave velocity structure was constrained to use V_p/V_s of 1.73.

The map area imaged, shot locations, and the initial epicentral locations for events used during inversion are shown (Figure 70). The latitude and longitude of the local coordinate system origin is $60^{\circ} 29' 20.4''$ N, $152^{\circ} 45' 11.4''$ W.

All models shown are based on a uniform node spacing of 0.4 km along all axes. We began with a 1.6 km spacing, and gradually reduced the spacing in increments of 0.2 km. During node-spacing reduction, velocities from the more coarsely spaced model were interpolated to initialize the finely spaced model. Smoothing coefficients of 0.3 for P wave perturbations and 2.5 for V_p/V_s perturbations were used. CDI constraints were also used to keep the model in the neighborhood of the well-resolved one-dimensional model. We used a 75% background fraction with V_p and V_s model parameter standard deviations given by 3.33 km s^{-1} and 1.4, respectively. Anomalous population standard deviations for V_p and V_p/V_s were 5 km s^{-1} and 2.0, respectively. All *a priori* information was interpretively selected during a series of inversions to balance prediction error against apparent model variance. The weighted RMS residual was reduced by 67% to 0.065 s.

Horizontal slices of P , S , and $\frac{V_p}{V_s}$ structure are shown for various depths (Figure 54 - Figure 62). Very few earthquakes occurred in the shallow section, and most the shallow slices only resolve a small fraction of the objective volume. The displays are presented in

terms of percent perturbation away from the one-dimensional model. This approach permits use of a uniform scale in the displays, but does not facilitate interpretation of absolute velocities. Final epicentral estimates are not substantially different from the initial locations, and no display of the final epicenters is presented.

Spurr

The absolute one-dimensional velocity structure at Spurr is less well resolved than at Redoubt, due to the lack of shot information and the smaller permanent network. Our initial estimate of the one-dimensional *P*-wave velocity structure is based on a smoothed perturbation of the previous one-dimensional layered structure used by the Alaska Volcano Observatory (Figure 53). A minimal 1D inverse was then derived and used as the reference model for the 3D inversions. Anomalies must be interpreted as relatively fast or slow regions with respect to the 1D inverse, which is not unique. We do not believe that our efforts to produce a shear wave velocity model produced reliable estimates, and no *S* models are presented here.

The map area imaged and the location of initial epicentral estimates considered for use during inversion is shown in Figure 63. The latitude and longitude of the local coordinate system origin is $61^{\circ} 15' 57.3121''$ N, $152^{\circ} 14' 17.48806''$ W. The 3D models are based on a 0.5 km grid-node spacing. We began with a 2.5 km spacing, and gradually reduced the spacing in increments of 0.5 km. During node-spacing reduction, velocity information from the more coarsely spaced model was interpolated to initialize the more finely spaced model. A smoothing coefficient of 0.25 was used. CDI

constraints were also used to keep the model in the neighborhood of the minimal one-dimensional inverse. We used a 75% background fraction. The background and anomalous population standard deviations used are 4.0 km s^{-1} and 6.0 km s^{-1} , respectively. All *a priori* information was selected during a series of inversions and chosen to balance prediction error against apparent model variance. The weighted RMS residual was reduced to 0.075 s, a reduction of 45%.

Horizontal slices of *P* wave velocity structure are shown for various depths (Figure 65 - Figure 67). Final earthquake epicenter estimates within 0.5 km of the horizontal slices are overlain. The displays are presented in terms of percent perturbation away from the one-dimensional model. Vertical cross-sections between A-A' and B-B' of V_p , V_s , and V_p/V_s perturbations are shown in Figure 68. Earthquakes within 0.5 kilometer of the vertical section are projected normally into the cross section, are shown as open circles. A map of the relocated epicenters actually used during inversion is shown in Figure 69.

Augustine

The Augustine shot experiments provide absolute velocity information for the portion of the volcanic edifice rising above sea level. The map area imaged is shown in Figure 70. The latitude and longitude of the local coordinate system origin is $59^{\circ} 21' 30.0'' \text{ N}$, $153^{\circ} 25' 30.0'' \text{ W}$. A larger scale figure of the model area imaged is also shown (Figure 71). Our estimate of the one-dimensional P-wave velocity structure is based on a 1D inverse using both 1977 and 1995 shot data. This model is shown together with the one-

dimensional layered structure currently used by the Alaska Volcano Observatory for routine earthquake location estimation (Figure 72). Observed traveltimes from the shots indicate a high degree of small-scale heterogeneity. Both 2D and 3D velocity models based upon manual interpretation of the 1977 shot data have been previously proposed (Pearson, 1977; Kienle et al., 1979) (Figure 73).

All models shown are based on a uniform node spacing of 0.25 km along all axes. We began with 1.0 km spacing, and gradually reduced the spacing in increments of 0.25 km. During node-spacing reduction, velocities from the more coarsely spaced model were interpolated to initialize the finely spaced model. A smoothing coefficient of 0.25 for was used. CDI constraints were also used to keep the model in the neighborhood of the minimal one-dimensional inverse. We used a 75% background fraction. The background and anomalous population standard deviations used are 4.0 km s^{-1} and 6.0 km s^{-1} , respectively. All *a priori* information was selected during a series of inversions and chosen balance prediction error against apparent model variance. The weighted RMS residual was reduced to 0.073 s, which represents a reduction of 28%. This is a fairly small reduction in overall prediction error, but efforts to reduce the magnitude of the constraints selected resulted in very large and rapidly varying (spatially) parameter estimates. This is partially attributable to the small number of well picked observations typically reported with Augustine earthquakes (usually 5-8 phases), but also may indicate heterogeneity at a scale smaller than can be described with the grid-node spacing selected.

Horizontal slices of P wave velocity are shown for various depths(Figure 74 - Figure 76). Relocated earthquake epicenters within 0.25 km of the horizontal slices are overlain. The displays are presented in terms of percent perturbation away from the one-dimensional model.

DISCUSSION

The velocity models presented show that most of the traveltimes residual for each location investigated can be attributed to relatively large amplitude ($> 10\%$ perturbation) near-surface features. In all three cases anomalous amplitudes decay to a maximum amplitude of a few percent perturbation at depths below 1-2 km below sea level. Although CDI tomography was designed to allow estimation of small high-amplitude features, we find the small anomalous amplitudes at depth to be somewhat surprising. This is particularly true in light of the fact that smoothing constraints were used in conjunction with CDI. There are two possible explanations for the anomalies estimated: 1) Shallow processes strongly influence seismic velocities near these volcanoes, and the deeper subsurface is fairly homogeneous; or 2) CDI constraints have accounted for traveltimes residuals actually associated with poorly resolved deep anomalies through additional and inappropriate perturbation of unquestionably anomalous shallow regions. Since the question hinges on near surface velocity resolution, the issue could be resolved with a dense recording array and a large number of shots. Until such data are available, the correct interpretation is open for discussion.

In all cases, the largest anomalies are consistent with other geological and geophysical observations. A brief review of the largest and most reliably imaged features for each volcano follows.

Redoubt

The largest anomalous features imaged at Redoubt are most apparent on the horizontal slice displays of V_p/V_s at levels of -0.4 km to 1.2 km above sea level. Three anomalous volumes with V_p/V_s values of $1.6 - 1.65$ are apparent. The largest amplitude anomaly occurs at local coordinate system coordinates of $(x = -2 \text{ km}, y = 2.5 \text{ km})$ (Figure 61). Similar values of V_p/V_s have been reported at Redoubt using smoothness based tomographic imaging with separate estimation of P and S velocities (Benz et al., 1996). Benz et al. (1996) interpreted the regions of low V_p/V_s as zones of extensive micro-cracking and hydrothermal circulation, both of which can markedly decrease the bulk-modulus or P -wave velocity. This interpretation is consistent with the physical model of the eruptive sequence proposed by Chouet et al. (1994). Their model specifies a shallow structure with physical properties dominated by hydrothermal circulation. The strong V_p/V_s anomalies are confined to regions above two kilometers below sea level. This might be interpreted as the depth at which micro-cracks are closed by the lithostatic load (Christensen, 1979).

Directly beneath the most anomalous region of V_p/V_s is narrow and prominent zone of elevated relative V_s . Most of the raypaths in this area travel almost vertically, and this

feature might be attributed to raypath streaking downward from the region of relatively high V_s .

We do not find any indication of a large low V_s zone indicative of a magma chamber at any point in the imaged volume. No large amplitude anomalies of either V_p or V_s are found at depth. In light of the demonstrated ability of CDI to successfully concentrate signal associated with small discrete anomalies into a single large amplitude region, our null result suggests that no significant quantity of magma is currently stored in the shallow crust.

Spurr

CDI tomography has successfully concentrated most of the anomalous signal at Spurr into two fairly strong anomalies. The largest and most easily interpreted anomaly is a region of low V_p (-14%) centered on a point three kilometers east and 4.5 kilometers south of Crater Peak (Figure 65 - Figure 67). The anomaly extends from a depth of one kilometer below sea level to a level of one kilometer above sea level (Figure 68). The structure dips slightly toward the northwest, and correlates with a previously identified region of low resistivity, elevated mercury and helium levels, and anomalous self-potential (Turner and Wescott, 1986; Wescott et al., 1988). They interpreted the region as one of three active hydrothermal systems along the southern flank of Mt. Spurr. The low V_p values in the zone might be attributable to a region of hydrothermal alteration in and around the hydrothermal system. No earthquake epicenters are located in the region, and no seismicity beneath the region preceded the 1992 eruption of Mt. Spurr. This

suggests that the zone was not actively involved or significantly disturbed by eruption related processes. The other hydrothermal systems reported are in regions of poor raypath coverage and are not imaged as low velocity zones.

The second anomaly at Spurr is a region of high velocity directly beneath the vent of Crater Peak. Anomalous values of approximately +5% extend from sea level to the surface. This feature might be an artifact caused by use of a reference model that is too slow. Manual perturbation of the initial velocity field toward higher velocities consistent with those estimated in the anomalous zone does indeed suppress the anomaly, but causes an 8% increase in initial weighted RMS prediction error. We therefore accept the slower minimal 1D inverse, and present the images shown here. The question of what the zone might be, if it is even a physical reality, is problematic. Mt. Spurr magmas are expected to be compositionally similar to each other, and the hypothesis of a more mafic intrusion or dike plexus is inconsistent with the geological model adopted by AVO. We refrain from further speculation about the nature of the anomaly, but suggest that the question of anomaly existence could be answered through a relatively small shot experiment.

No low velocity zones that might be interpreted as a magma chamber or region of significant magma storage were imaged in the upper nine kilometers of the crust.

Augustine

Most of Augustine Island is covered by a pyroclastic apron and repeated debris avalanche deposits associated with various scales of dome collapse (Waitt and Beget, 1996). Refraction experiments indicate a very low velocity ($V_p = 1.8$) for the pyroclastic

apron (Pearson, 1977). The core of the volcano is expected to have a very high V_p relative to the outer skin. A sequence of endogenous domes occupies the summit region (Kienle and Forbes, 1976).

The tomograms of Augustine do not provide the detail required to address questions of dome structure, but do confirm the existence of a very high speed ($> 50\%$ perturbation) core relative to the surface material (Figure 74). Our estimate of outer skin velocity is consistent with previous hypotheses, but the estimate of central core velocity is somewhat lower than that proposed by previous investigators (Kienle et al. 1979). The previous interpretation was based on the assumption of a symmetric structure and a hypothetical core radius. Our images suggest that the high velocity core extends from a point near the 1986 vent northward to the flank of the cone. Our relatively slow velocity estimates toward the south could be accurate, although the region is not well illuminated and the low velocities might be a result of a more spatially limited region of fracturing and hydrothermal alteration immediately beneath the south rim stations.

CONCLUSIONS

Seismic geotomography is used to generate non-unique estimates of parameter perturbations that bring an initial model toward a higher degree of consistency with observations. The ability to accurately model raypaths and partial derivatives of traveltimes is required to correctly estimate optimal velocity parameter perturbations. Seismic tomography applied to volcanic systems requires the ability to accurately trace rays through extremely heterogeneous models with large velocity contrasts and in the

presence of rugged topography. Accurate estimation of optimal hypocenter locations and sufficient data redundancy is required. An infinite number of optimal models exist, and *a priori* information must be added to the problem to select a point solution. The model selected as a 'solution' is highly dependent upon the type of *a priori* information used. Shot information can provide absolute velocity information that reduces non-uniqueness in the relatively shallow subsurface.

An implementation that addresses the challenges of volcano tomography has been developed. Use of full non-linear hypocenter location estimation in conjunction with robust 3D raytracing and CDI constraints indicate that it is possible to explain most of the observed traveltimes residual at Redoubt, Mt. Spurr, and Augustine volcanoes with large amplitude, shallow anomalies. The data do not require a strongly heterogeneous deeper subsurface. The more correct interpretation cannot be identified without absolute velocity information throughout the uppermost portion of the volcanic edifice.

ACKNOWLEDGEMENTS

Projects were funded in part by NSF grants EAR-94-05471, EAR-95-06379, and the Alaska Volcano Observatory. Scholarships were provided by the Alaska Geophysical Society through the SEG Scholarship Foundation .

Computational resources used for this investigation were provided by the University of Alaska Fairbanks Geophysical Institute (Seismology Laboratory), the Alaska Volcano Observatory, the U.S. Geological Survey (Anchorage, AK), and Applied Digital Technologies.

Phil Dawson of the U.S.G.S. provided data acquired with temporary array at Redoubt Volcano. The 1977 Augustine shot experiment was conducted by Juergen Kienle, C. F. Pearson, and Doug Lalla. Sandia Laboratories provided instrumentation. Juergen Kienle archived information from the 1977 investigation and provided it for use in this study. John Lahr provided much information about the current art of hypocenter location as well as assistance with data retrieval. The 1995 Augustine field experiment was conducted by a group of graduate student volunteers and summer interns that deserve much credit. Finn Michelsen, as an OYO corporation representative, made a large multi-channel data acquisition system available for the Augustine during a time window when weather prevented deployment. John Power provided assistance and historical information concerning both field programs and data at Augustine.

The opinions presented here are those of the authors alone and do not necessarily reflect the opinion of the Alaska Volcano Observatory or the University of Alaska Fairbanks Geophysical Institute.

REFERENCES

- Abers, G. A., 1994, Three-dimensional inversion of regional P and S arrival times in the East Aleutians and sources of subduction zone gravity highs: *J. Geophys. Res.*, **89**, 6079-9094.
- Aki, K., and Richards, P., 1980, *Quantitative seismology: Theory and methods*: W. H. Freeman & Co.

- Benz, H. M., Chouet, B. A., Dawson, P. B., Lahr, J. C., Page, R. A., and Hole, J. A., 1996, Three-dimensional *P* and *S* wave velocity structure of Redoubt Volcano, Alaska: *J. Geophys. Res.*, **101**, 8111-8128.
- Berryman, J. G., 1990, Lecture notes on nonlinear inversion and tomography, Lawrence Livermore National Laboratory Report, UCRL-LR-105358, 1990.
- Carrion, P., 1991, Dual tomography for imaging complex structures: *Geophysics*, **56**, 1395-1404.
- Carrion, P. M., Puledda, S., and Comelli, P., 1997, Gibbs statistics in crosswell and reflection tomography: Inversion of noisy data: *Geophysics*, **62**, 1208-1213.
- Chouet, B. A., Page, R. A., Stephens, C. D., Lahr, J. C., and Power, J. A., 1994, Precursory swarms of long-period events at Redoubt Volcano (1989-1990), Alaska: Their origin and use as a forecasting tool, *J. Volcan. Geotherm. Res.*, **62**, 95-136.
- Cerveny, V., 1985, The application of ray tracing to the numerical modelling of seismic wave fields in complex structures, in *Handbook of geophysical exploration*, Section 1, Seismic exploration, Vol. 15A: Geophysical Press, 1-119.
- Cerveny, V., and Psencik, I., 1983, Gaussian beams and paraxial ray approximation in three-dimensional elastic inhomogeneous media, *J. Geophys.*, **53**, 1-15.
- Christensen, N. I., 1979, Compressional wave velocities in rocks at high temperatures and pressures, critical thermal gradients, and crustal low-velocity zones: *J. Geophys. Res.*, **84**, 6849-6857.
- Clippard, J. D., and Christensen, D. H., 1995, Composite distribution inversion applied to crosshole tomography: *Geophysics*, **60**, 1283-1294.
- Dines, K. A., and Lytle, R. J., 1979, Computerized geophysical tomography, *Proc. IEEE*, **67**, 1065-1073.

- Geoltrain, S. and Brac, J., 1993, Can we image complex structures with first arrival traveltimes?: *Geophysics* **58**, 564-575.
- Gilbert, P., 1972, Iterative methods for the three-dimensional reconstruction of an object from projections: *J. Theor. Biol.*, **36**, 105-117.
- Gill, P. E., Murray, W., and Wright, M. H., 1981, *Practical optimization*: Academic Press.
- Golub, G., and Reinsch, C., 1970, Singular value decomposition and least-squares solutions: *Numer. Math.* **14**, 403-420.
- Gray, S. H. and May W. P., 1994, Kirchhoff migration using eikonal equation traveltimes: *Geophysics*, **59**, 810-817.
- Grechka, V. Y., and McMechan, G. A., 1996, 3-D two-point ray tracing for heterogeneous, weakly transversely isotropic media: *Geophysics*, **61**, 1883-1894.
- Hanyga, A., 1988, Numerical methods for tracing rays and wavefronts, in Doornbos, D. J., Ed., *Seismological algorithms: Computational methods and computer programs*: Academic Press.
- Hestenes, M., and Stiefel, E., 1952, Methods of conjugate gradients for solving linear systems: *Nat. Bur. Standards J. Res.*, **49**, 409-436.
- Hounsfield, G. N., 1973, Method and apparatus for measuring X or gamma radiation absorption or transmission at plural angles and analyzing the data: U.S. Patent 3 778 614.
- Humphreys, E., and Clayton, R. W., 1988, Adaptation of back projection tomography to seismic travel time problems: *J. Geophys. Res.*, **93**, 1073-1085.
- Hyndman, D. W., and Harris, J. M., 1996, Traveltime inversion for the geometry of aquifer lithologies: *Geophysics*, **61**, 1728-1737.

- Hyndman, D. W., Harris, J. M., and Gorelick, S. M., 1994, Coupled seismic and tracer test inversion for aquifer property characterization: *Water Resources Res.*, **30**, 1965-1977.
- Ivansson, S., 1986, Seismic borehole tomography - Theory and computational methods: *Proc. IEEE*, **74**, 328-338.
- Julian, B. R., and Gubbins, D., 1977, Three dimensional seismic ray tracing: *J. Geophys.*, **43**, 95-114.
- Kaczmarz, S., 1937, AngenŠherte Auflšsung von Systemen linearer Gleichungen: *Bull. Acad. Polon. Sci. Lett.*, **A**, 355-357.
- Keller, H. B., and Perozzi, D. S., 1983, Fast seismic ray tracing: *SIAM J. Appl. Math.*, **43**, 981-992.
- Kienle, J. and Forbes, R. B., 1976, Augustine-evolution of a volcano: *Geophysical Institute, Annual Report 1975-1976*, R-275, 26-48.
- Kienle, J., Lalla, D. J., Pearson, C. F., and Barrett, S. A., 1979, Search for shallow magma accumulations at Augustine Volcano: *Final Report to U. S. Department of Energy, contract AT(45-1)-2229 task agreement 8*.
- Kissling, E., Ellsworth, W. L., Eberhart-Phillips, D., and Kradolfer, U., 1994, Initial reference models in local earthquake tomography: *J. Geophys. Res.*, **99**, 19635-19646.
- Klimes, L., and Kvasnicka, M., 1994, 3-D network ray tracing: *Geophys. J. Int.*, **116**, 726-738.
- Lahr, J. C., 1980, HYPOELLIPSE/MULTICS: A computer program for determining local earthquake hypocentral parameters, magnitude, and first motion pattern: *U.S. Geol. Surv. Open-File Report 80-59*.

- Lahr, J. C., Chouet, B. A., Stephens, C. D., Power, J. A., and Page, R. A., 1994, Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989-1990 eruptions at Redoubt Volcano, Alaska, *J. Geotherm. Res.*, **62**, 137-152, 1994.
- Lanczos, C., 1961, *Linear differential operators*: Van Nostrand, New York, 1961.
- Langan, R. T., Kerche, I., and Cutler, R. T. 1985, Tracing of rays through heterogeneous media: An accurate and efficient procedure: *Geophysics*, **50**, 1456-1465.
- Lees, J. M., and Crosson, R. S., 1989, Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data: *J. Geophys. Res.*, **94**, 5716-5728.
- Lees, J. M., and Shalev, E., 1992, On the stability of P-wave tomography at Loma Prieta: A comparison of parameterizations, linear and nonlinear inversions: *Bull. Seis. Soc. Am.*, **82**, 1821-1839.
- Lines, L. R., and Treitel, S., 1984, A review of least-squares inversion and its application to geophysical problems: *Geophys. Prosp.*, **32**, 159-186.
- Lines, L. R., and LaFehr, T., 1989, Tomographic modeling of a cross-borehole data set: *Geophysics*, **54**, 1249-1257.
- Menke, W., 1984, *Geophysical data analysis: Discrete inverse theory*: Academic Press Inc.
- McNutt, S. R., 1996, Seismic monitoring and eruption forecasting of volcanoes: A review of the state-of-the-art and case histories. In *Monitoring and mitigation of volcano hazards*, R. Scarpa and R. I. Tilling eds. Springer-Verlag Berlin, 99-146.
- Menke, W., 1984, *Geophysical data analysis: Discrete inverse theory*: Academic Press Inc.
- Moser, T. J., 1991, Shortest path calculation of seismic rays: *Geophysics*, **56**, 59-67.

- Moser, T. J., van Eck, T., and Nolet, G., 1992, Hypocenter determination in strongly heterogeneous earth models using the shortest path method: *J. Geophys. Res.*, **97**, 6563-6572.
- Nichols, D. E., 1996, Maximum energy traveltimes calculated in the seismic frequency band: *Geophysics*, **61**, 253-263.
- Nolet, G., 1985, Solving or resolving inadequate and noisy tomographic systems: *J. Comp. Phys.*, **61**, 463-482.
- Paige, C. C., and Saunders, M. A., 1982, Algorithm 583, LSQR: Sparse linear equations and least-squares problems, *ACM Trans. Math. Soft.*, **8**, 195-209
- Pavlis, G. L., and Booker, J. R., 1980, The mixed discrete-continuous inverse problem: application to the simultaneous determination of hypocenters and velocity structure: *J. Geophys. Res.*, **85**, 4801-4810.
- Pearson, C. F., 1977, Seismic refraction study of Augustine Volcano: University of Alaska M.S. thesis, 131 p.
- Podvin, P., and Lecomte, I., 1991, Finite-difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools: *Geophys. J. Internat.*, **105**, 271-284.
- Polak, E., 1971, *Computational methods in optimization*: Academic Press.
- Prothero, W. A., Eickemeyer, J., and Taylor, W. J., 1988, A fast two-point raytracing algorithm using a simple step search method: *Bull. Seism. Soc. Am.*, **78**, 1190-1198.
- Sambridge, M. S., 1990, Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D: *Geophys. Journ. Int.*, **102**, 653-677.

- Sambridge, M. S., and Kennett, B. L. N., 1990, Boundary value ray tracing in heterogeneous medium: A simple and versatile algorithm: *Geophys. J. Internat.*, **101**, 157-168.
- Scales, J. A., 1987, Tomographic inversion via the conjugate gradient method: *Geophysics*, **52**, 179-185.
- Schneider, W. A., Ranzinger, K. A., Balch, A. H., and Kruse, C., 1992, A dynamic programming approach to first arrival traveltimes computation in media with arbitrarily distributed velocities, *Geophysics*, **57**, 39-50.
- Singh, R. P., and Singh, Y. P., 1991, RAYPT-- A new inversion technique for geotomographic data: *Geophysics*, **56**, 1215-1227.
- Snyder, J. P., 1982, Map Projections used by the U. S. Geological Survey: U.S. Geol. Surv. Bull. 1532.
- Stroustrup, B., 1991, The C++ programming language: AT&T Bell Telephone Laboratories, Inc.
- Stroustrup, B., and Ellis, M. A., 1994, The annotated C++ reference manual, AT&T Bell Laboratories, Inc.
- Tanabe, 1971, Projection method for solving a singular system of linear equations and its applications: *Numer. Math.*, **17**, 203-214.
- Tarantola, A., and Valette, B., 1982, Inverse Problems = quest for information, *J. Geophys.*, **50**, 159-170.
- Thurber, C. H., 1985, Nonlinear earthquake location: Theory and examples: *Bull. Seismol. Soc. Amer.*, **75**, 779-790.

- Turner, D. L., and Wescott, E. M., 1986, Summary and conclusion of the Mount Spurr Alaska geothermal assessment project: *in* Geothermal energy resource investigations at Mount Spurr Alaska, Turner, D. L., Wescott, E. M., ed., University of Alaska Geophysical Institute Report UAG R-308, 6.1-6.6.
- Um, J., and Thurber, C., 1987, A fast algorithm for two-point seismic ray tracing: *Bull. Seis. Soc. Am.*, **77**, 972-986.
- Vasco, D. W., Peterson J. E. Jr., and Major, E. L., 1996, Nonuniqueness in traveltime tomography: Ensemble inference and cluster analysis: *Geophysics*, **61**, 1209-1227.
- Vidale, J., 1988, Finite-difference calculation of traveltimes: *Bull. Seis. Soc. Am.*, **78**, 2062-2076.
- Vidale, J., 1990, Finite-difference calculation of traveltimes in three dimensions: *Geophysics*, **55**, 521-526.
- Villaseñor, A., Benz, H. M., De Luca, G., Filippi L., Patane, G., Scarpa, R., and Vinciguerra, S., 1997, Three-dimensional P-wave velocity structure of Mt. Etna, Italy: Submitted to *Geophys. Res. Lett.*
- Virieux, J., Farra, V., and Madariaga, R., 1988, Ray tracing in laterally heterogeneous media for earthquake location: *J. Geophys. Res.*, **93**, 6585-6599.
- Waite, R. B. and Begét, J. E., 1996, Provisional Geologic Map of Augustine Volcano, Alaska: U.S. Geol. Surv. Open-File Report 96-516.
- Wescott, E. M., Turner, D. L., Nye, C. J., Motyka, R. J., and Moore, P., 1988 Exploration for geothermal resources at Mt. Spurr, Alaska: *Geotherm. Resources Council Trans.*, **12**, 203-210.

Wielandt, E., 1987, On the validity of the ray approximation for interpreting delay times,
in Nolet, G., Ed., Seismic tomography with applications *in* global seismology and
exploration geophysics: D. Reidel Publ. Co., 85-98.

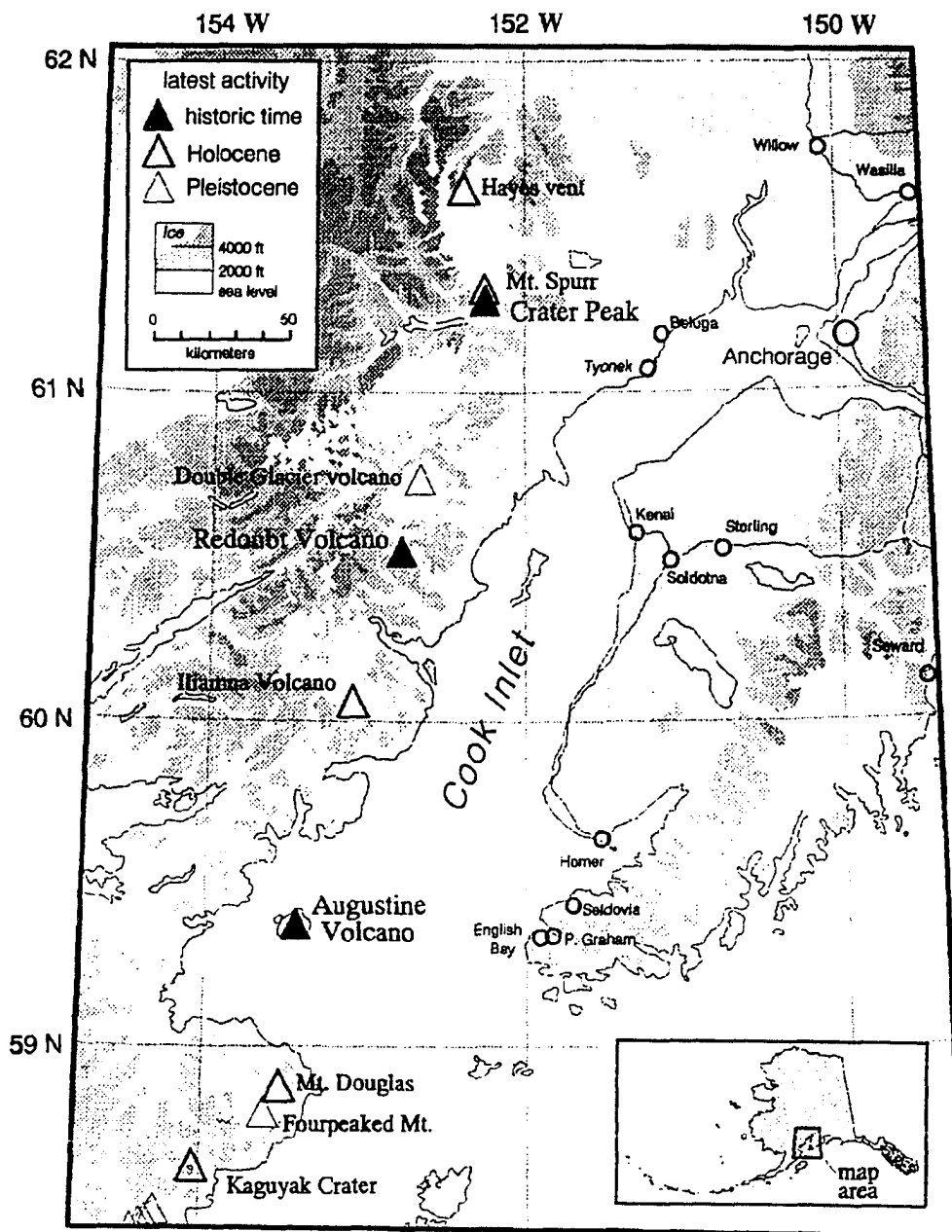


Figure 50: Location of Redoubt, Spurr, and Augustine volcanoes. Other Quaternary volcanoes in the Cook Inlet region of south-central Alaska are also shown.

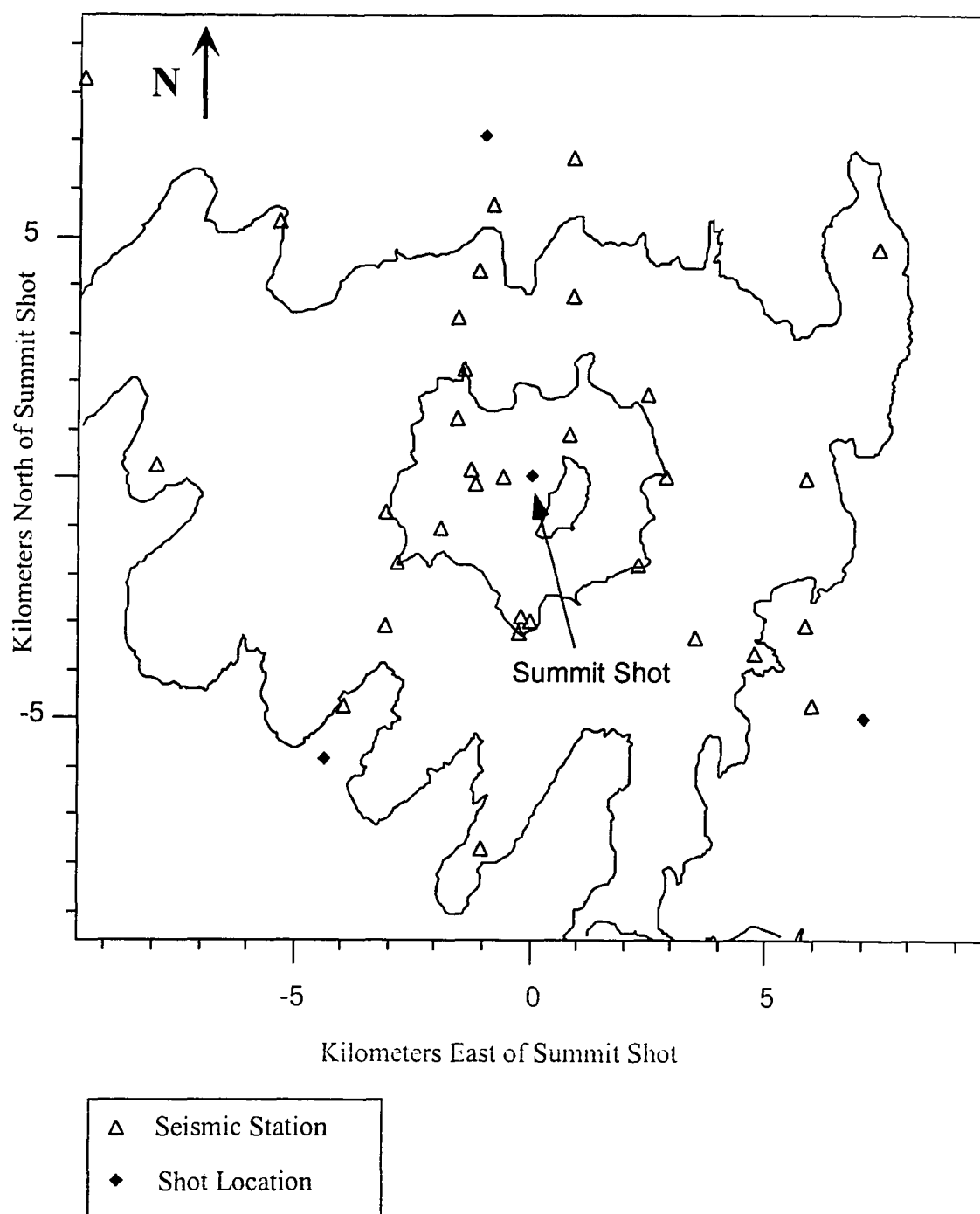


Figure 51: Redoubt Volcano study area. Both permanent and temporary seismic stations are shown as open triangles. Shots are shown as solid diamonds. Topographic contours at 900, 1830 and 2750 meters are shown for reference. The velocity models presented cover the entire shown.

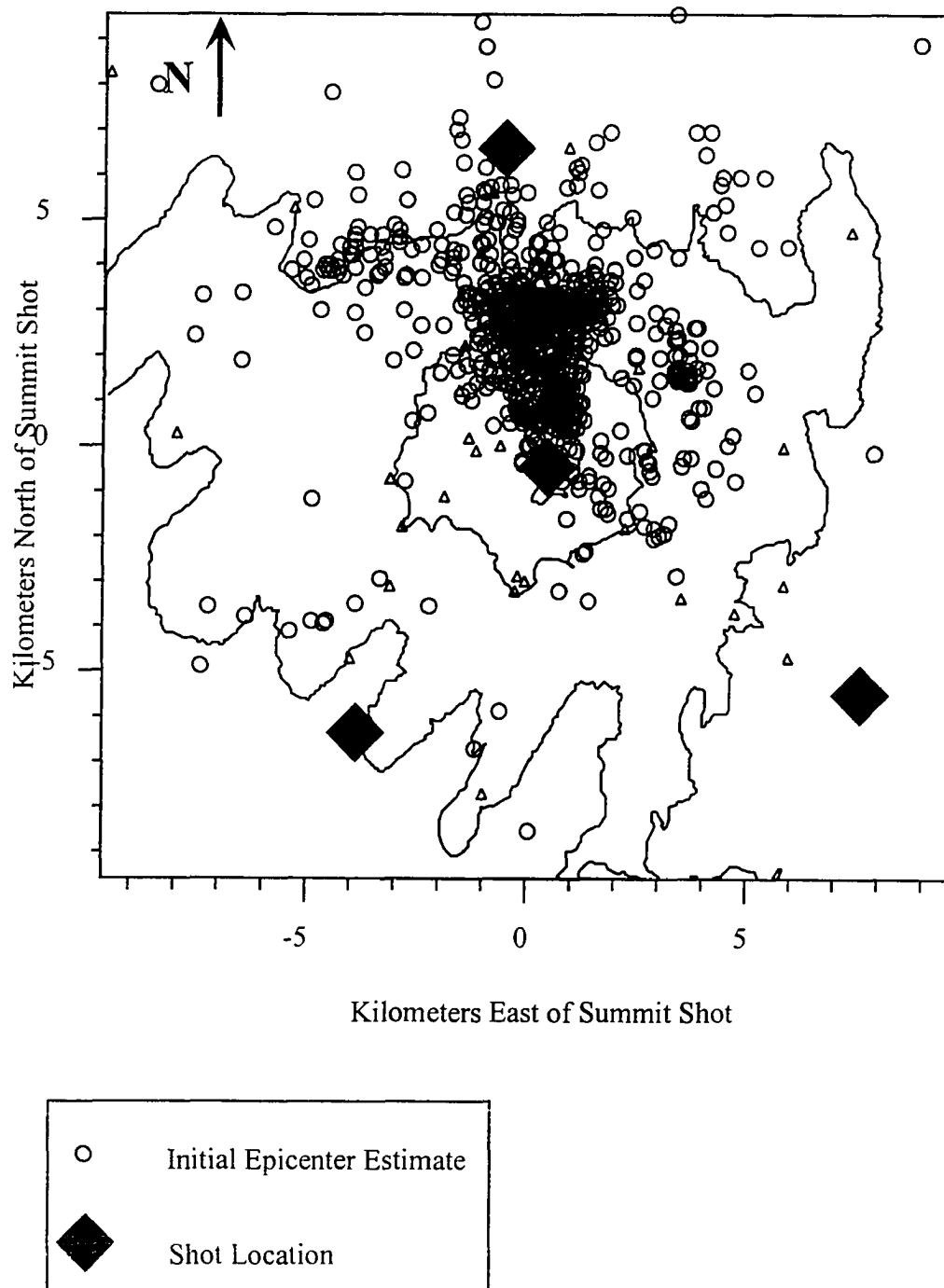


Figure 52: Initial epicenter estimates and shot locations used for tomographic imaging of Redoubt Volcano..

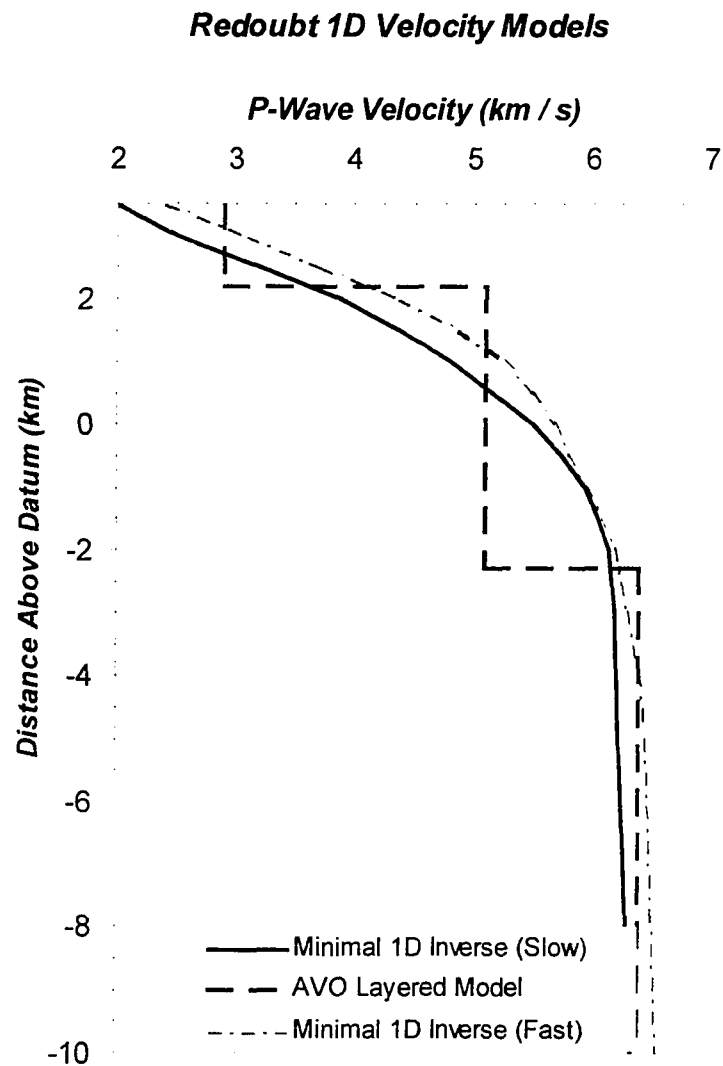


Figure 53: One dimensional Redoubt velocity models.

Tomographic images presented here use the faster minimal 1D inverse. The slower inverse was based on a slower initial model and satisfies the data equally well. The layered model was developed by Lahr et al. (1994) and is used by the Alaska Volcano Observatory for routine earthquake location estimation.

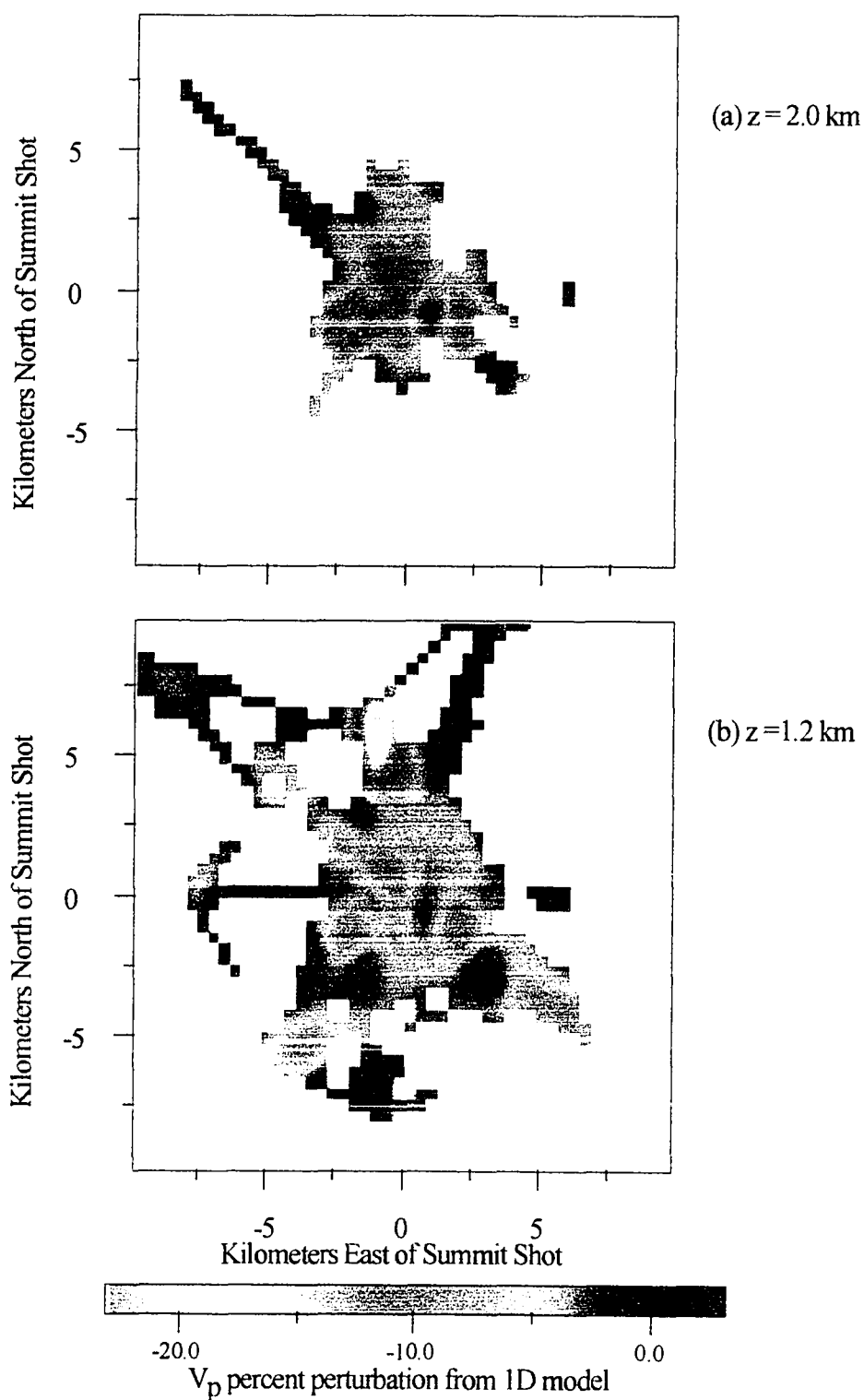


Figure 54: Horizontal slices of P wave velocity perturbation for Redoubt Volcano at 2 km and 1.2 km above sea level. a) 2 km above sea level; b) 1.2 km above sea level.

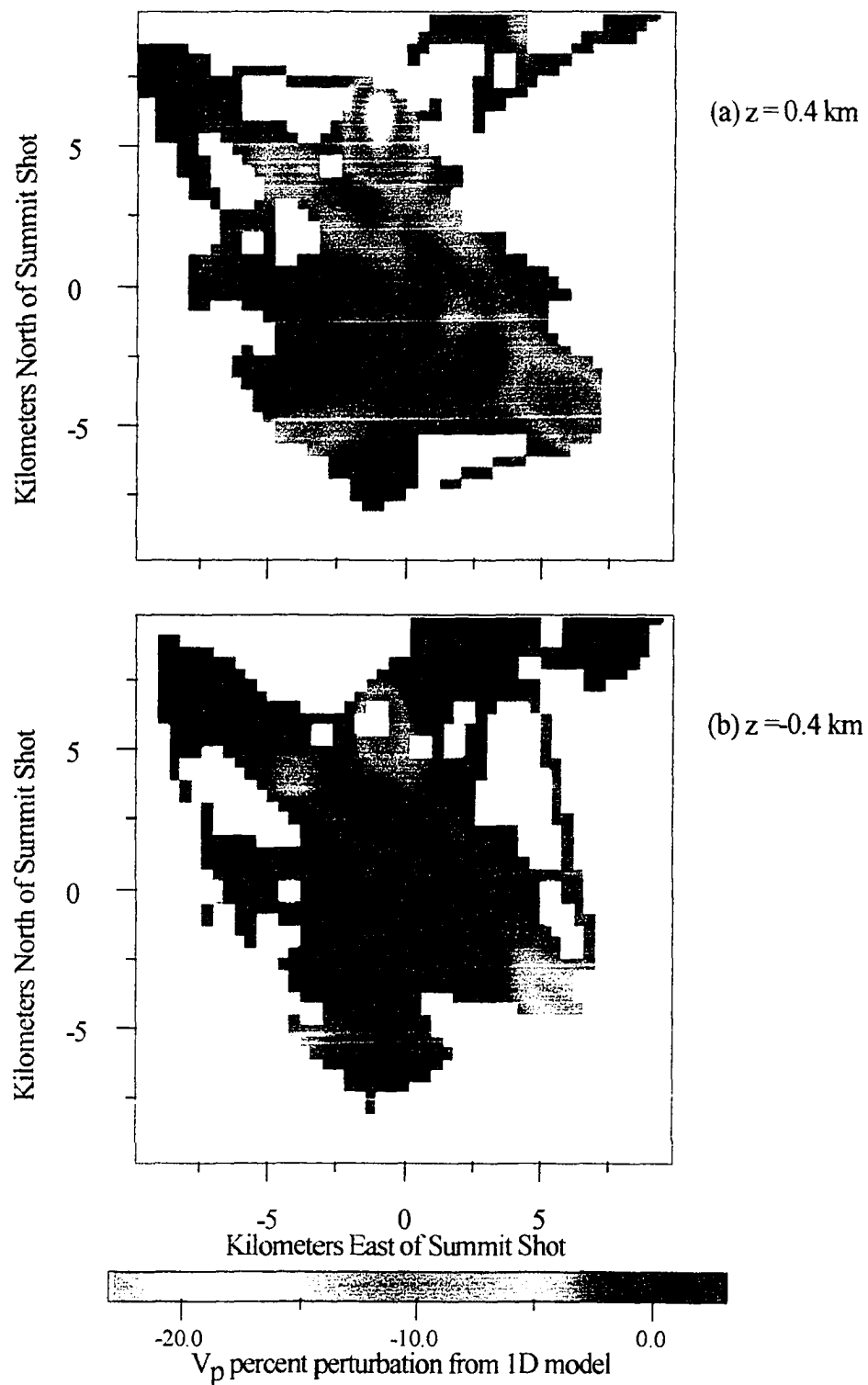


Figure 55: Horizontal slices of P wave velocity perturbation for Redoubt Volcano at 0.4 km above and 0.4 km below sea level. a) 0.4 km above sea level; b) 0.4 km below sea level.

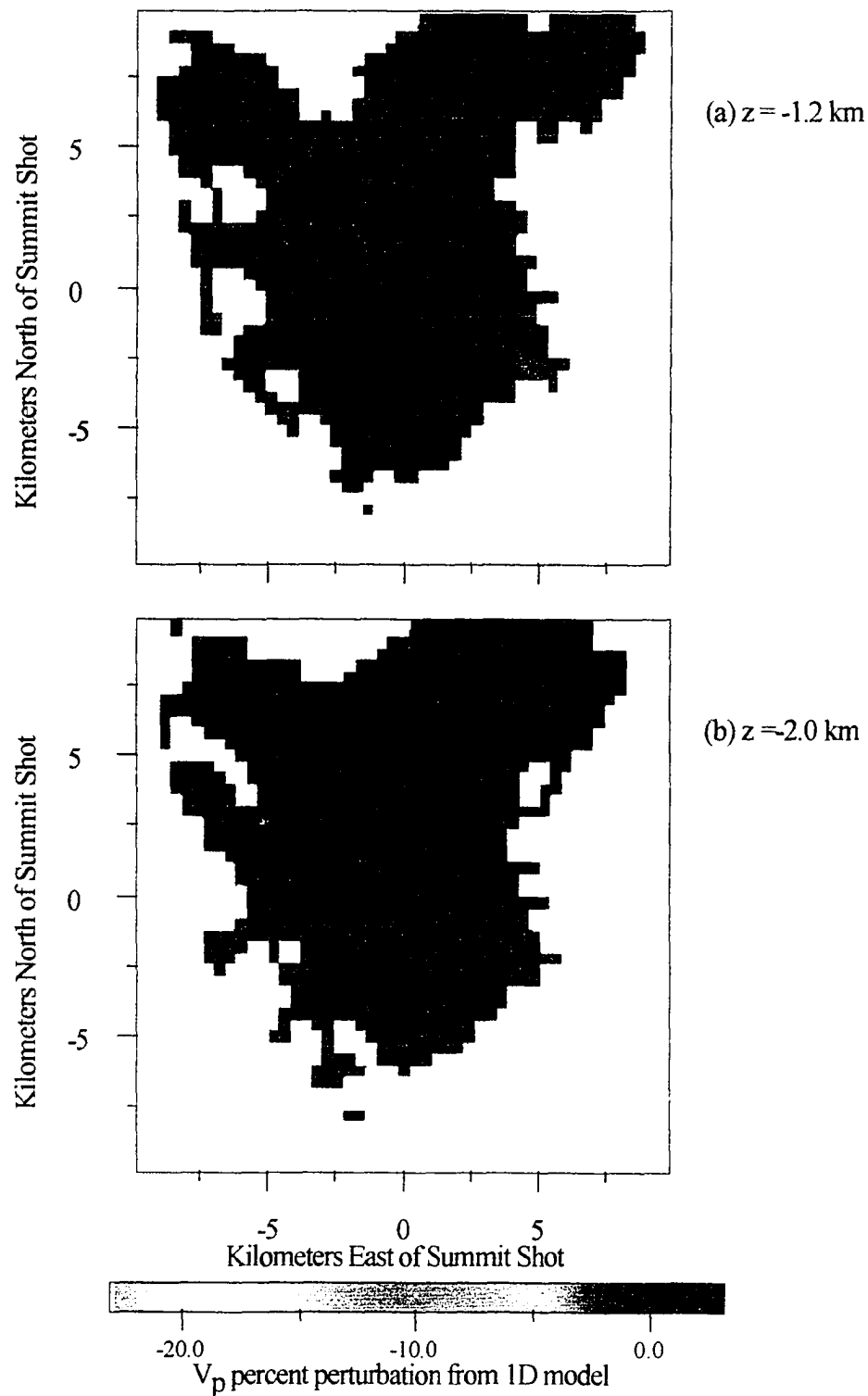


Figure 56: Horizontal slices of P wave velocity perturbation for Redoubt Volcano at 1.2 and 2 km below sea level. a) 1.2 km above sea level; b) 2.0 km below sea level.

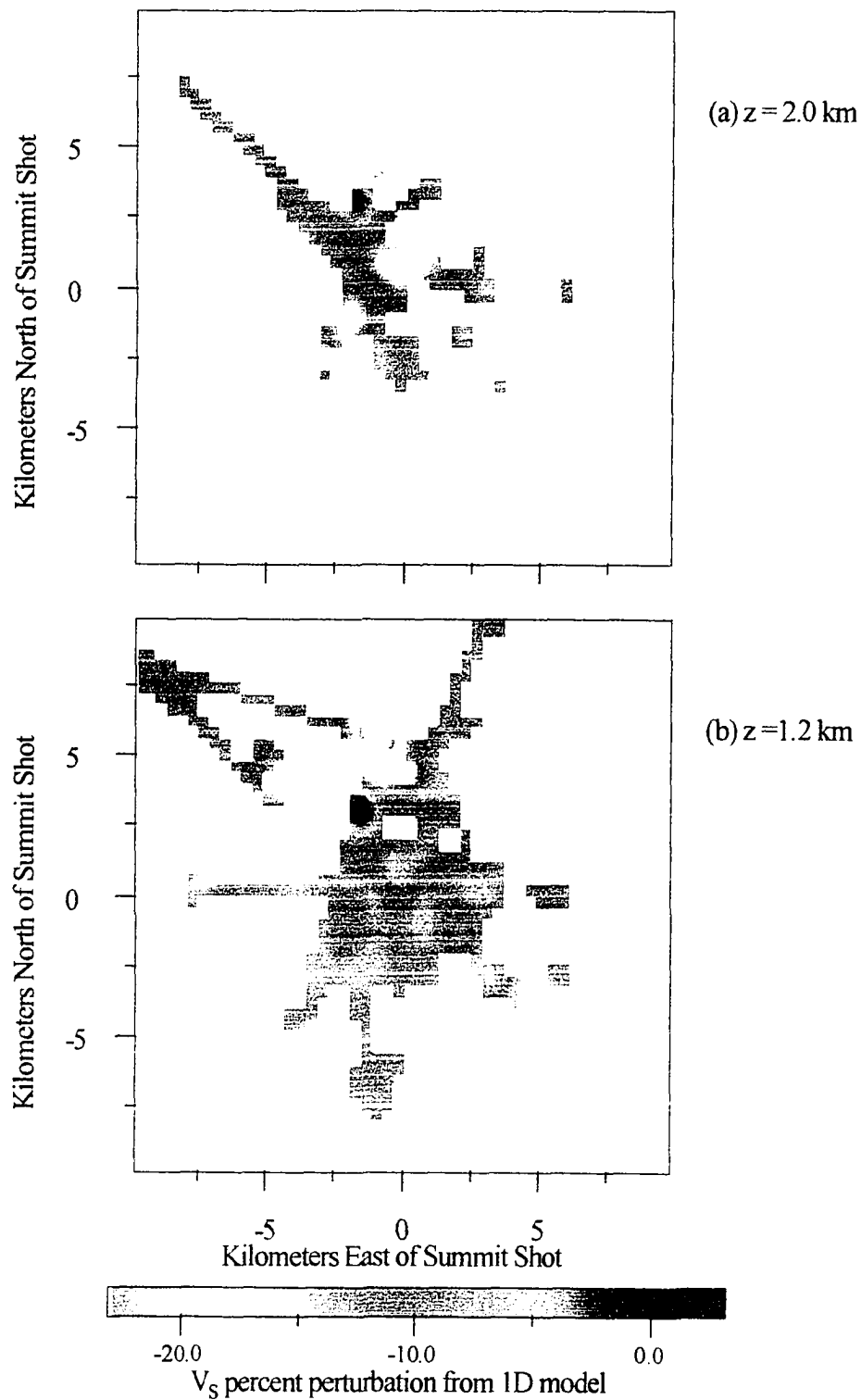


Figure 57: Horizontal slices of S wave velocity perturbation for Redoubt Volcano at 2 km and 1.2 km above sea level. a) 2 km above sea level; b) 1.2 km above sea level.

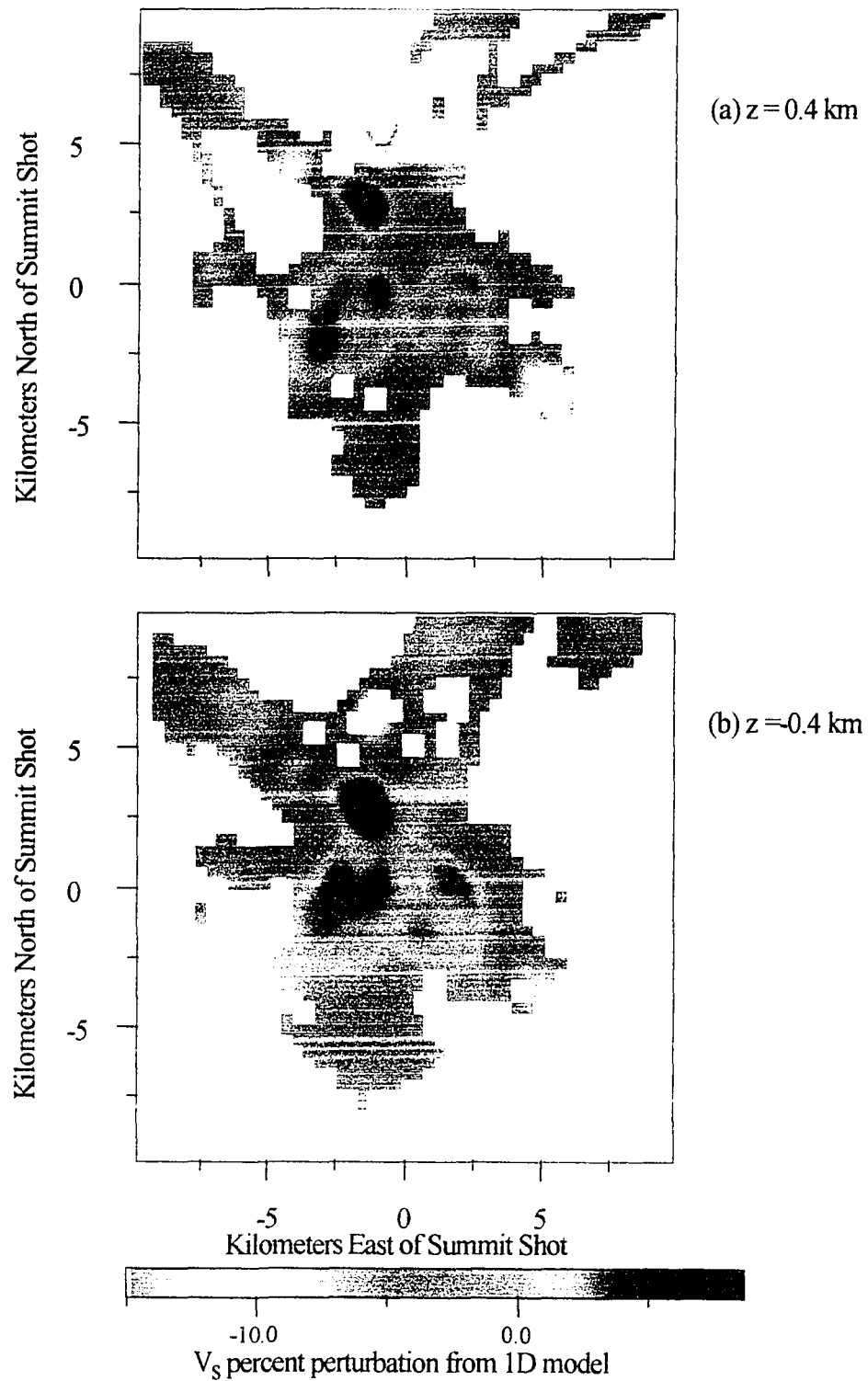


Figure 58: Horizontal slices of S wave velocity perturbation for Redoubt Volcano at 0.4 km above and 0.4 km below sea level. a) 0.4 km above sea level; b) 0.4 km below sea level.

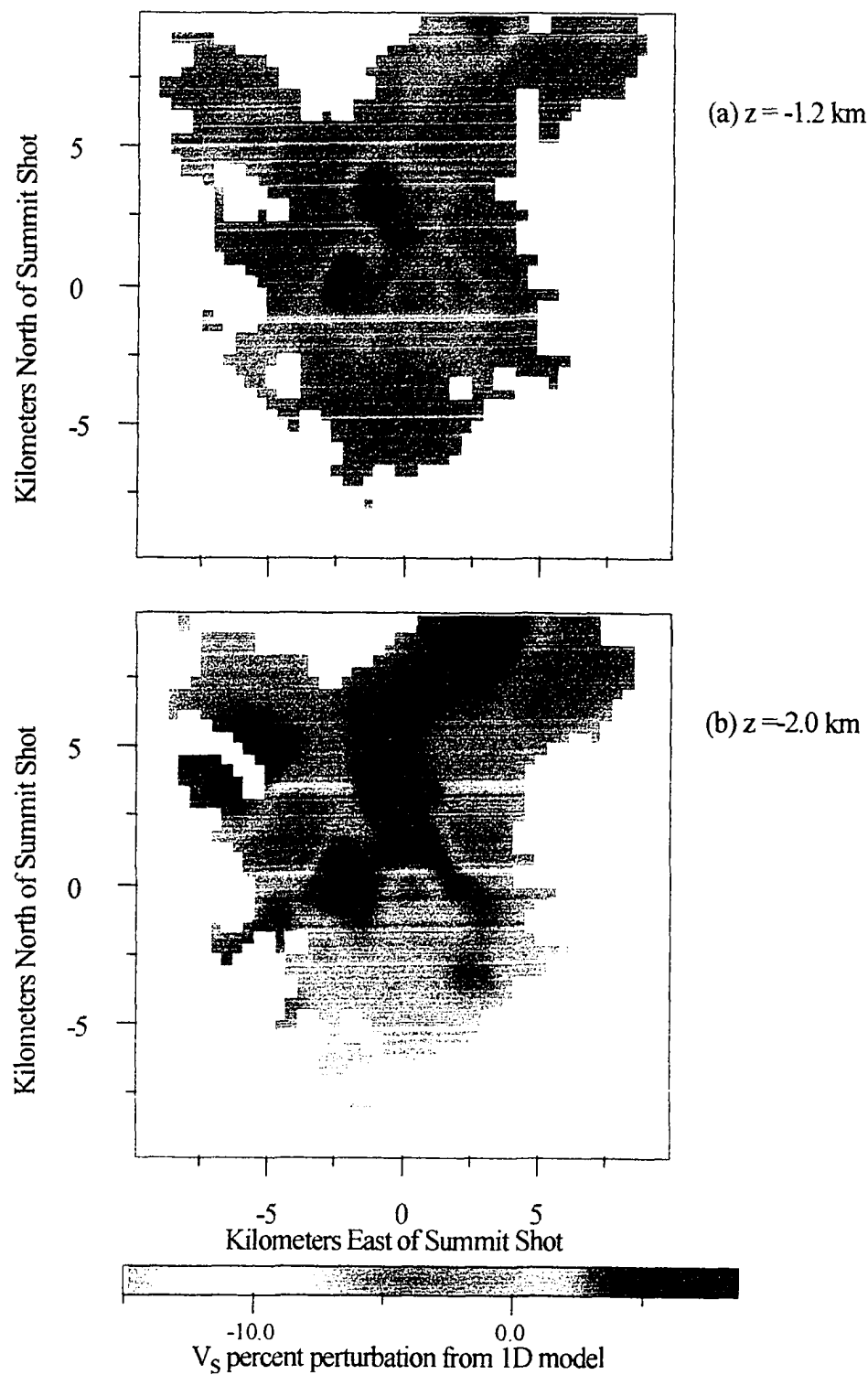


Figure 59: Horizontal slices of S wave velocity perturbation for Redoubt Volcano at 1.2 and 2 km below sea level. a) 1.2 km above sea level; b) 2.0 km below sea level.

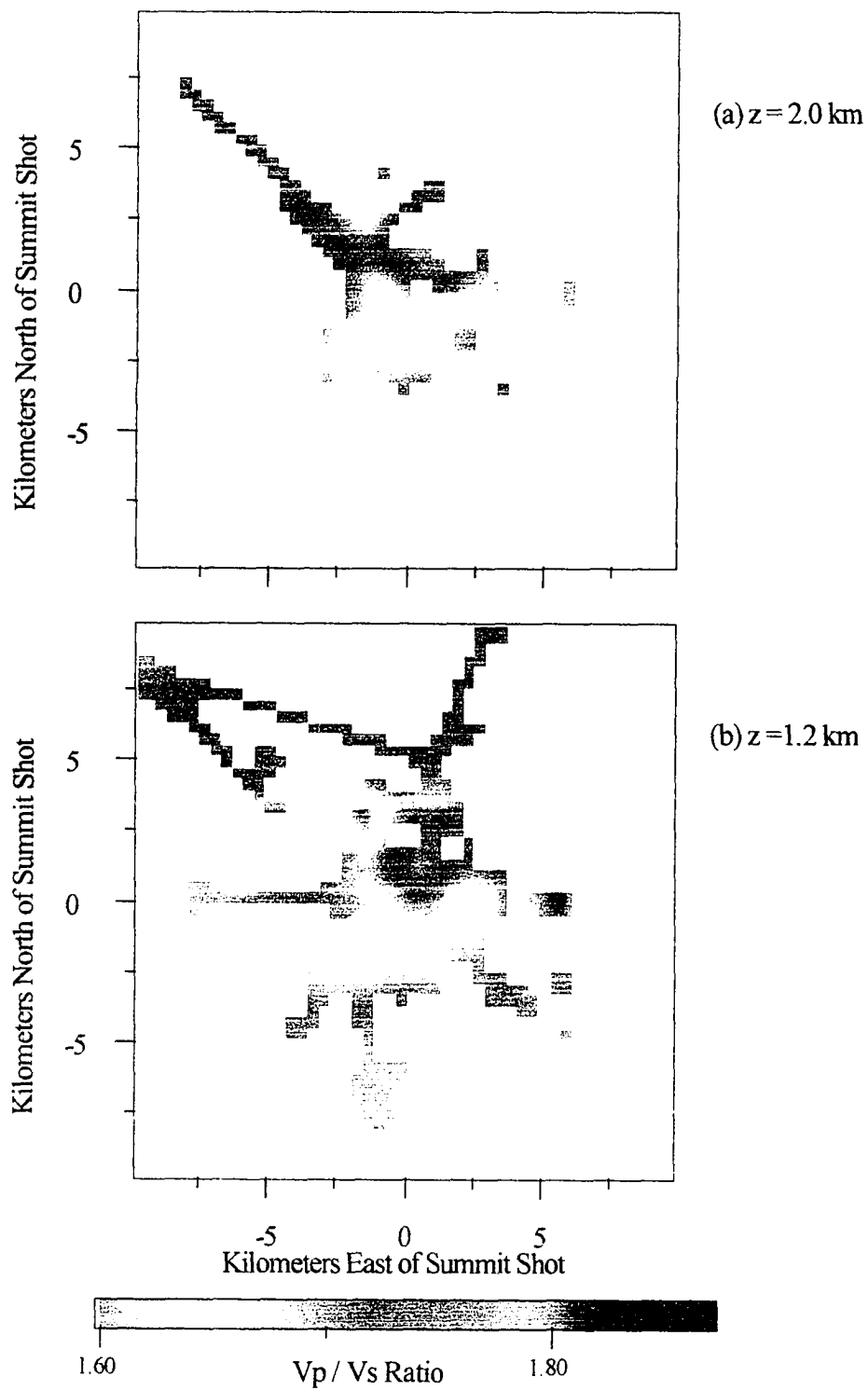


Figure 60: Horizontal slices of V_p / V_s estimate for Redoubt Volcano at 2 km and 1.2 km above sea level. a) 2 km above sea level; b) 1.2 km above sea level.

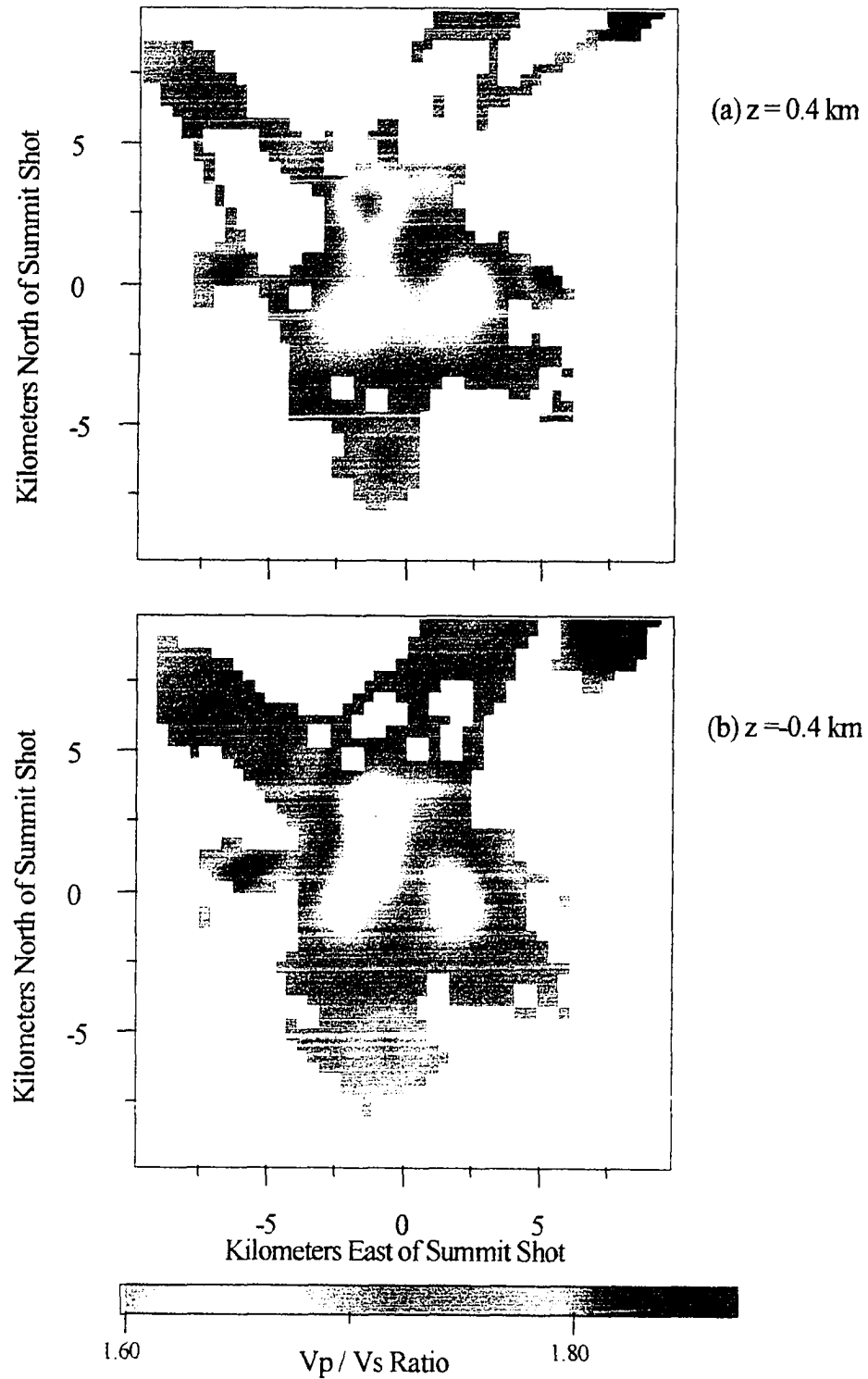


Figure 61: Horizontal slices of Vp / Vs estimate for Redoubt Volcano at 0.4 km above and 0.4 km below sea level. a) 0.4 km above sea level; b) 0.4 km below sea level.

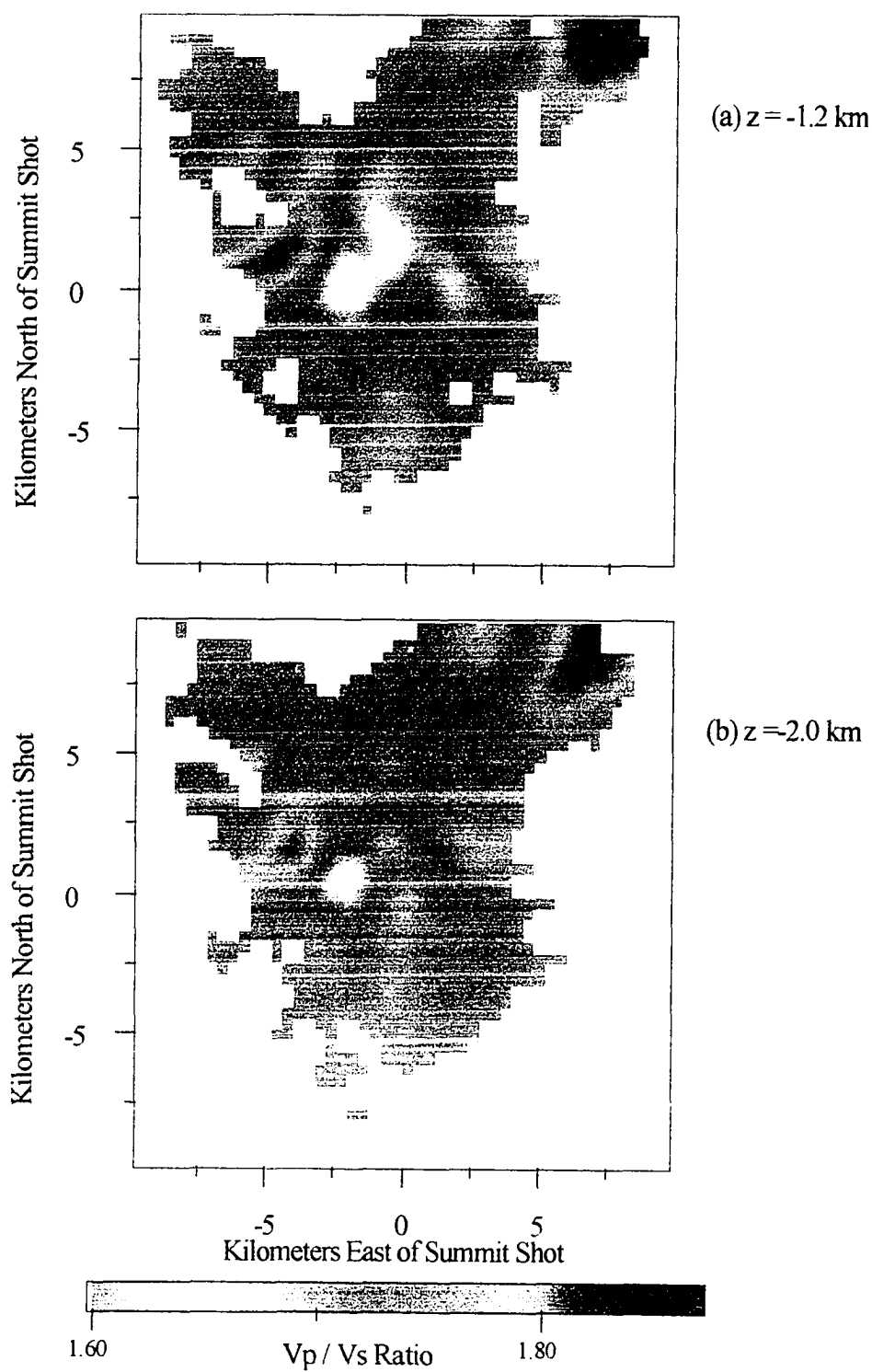


Figure 62: Horizontal slices of V_p/V_s estimate for Redoubt Volcano at 1.2 and 2 km below sea level. a) 1.2 km above sea level; b) 2.0 km below sea level.

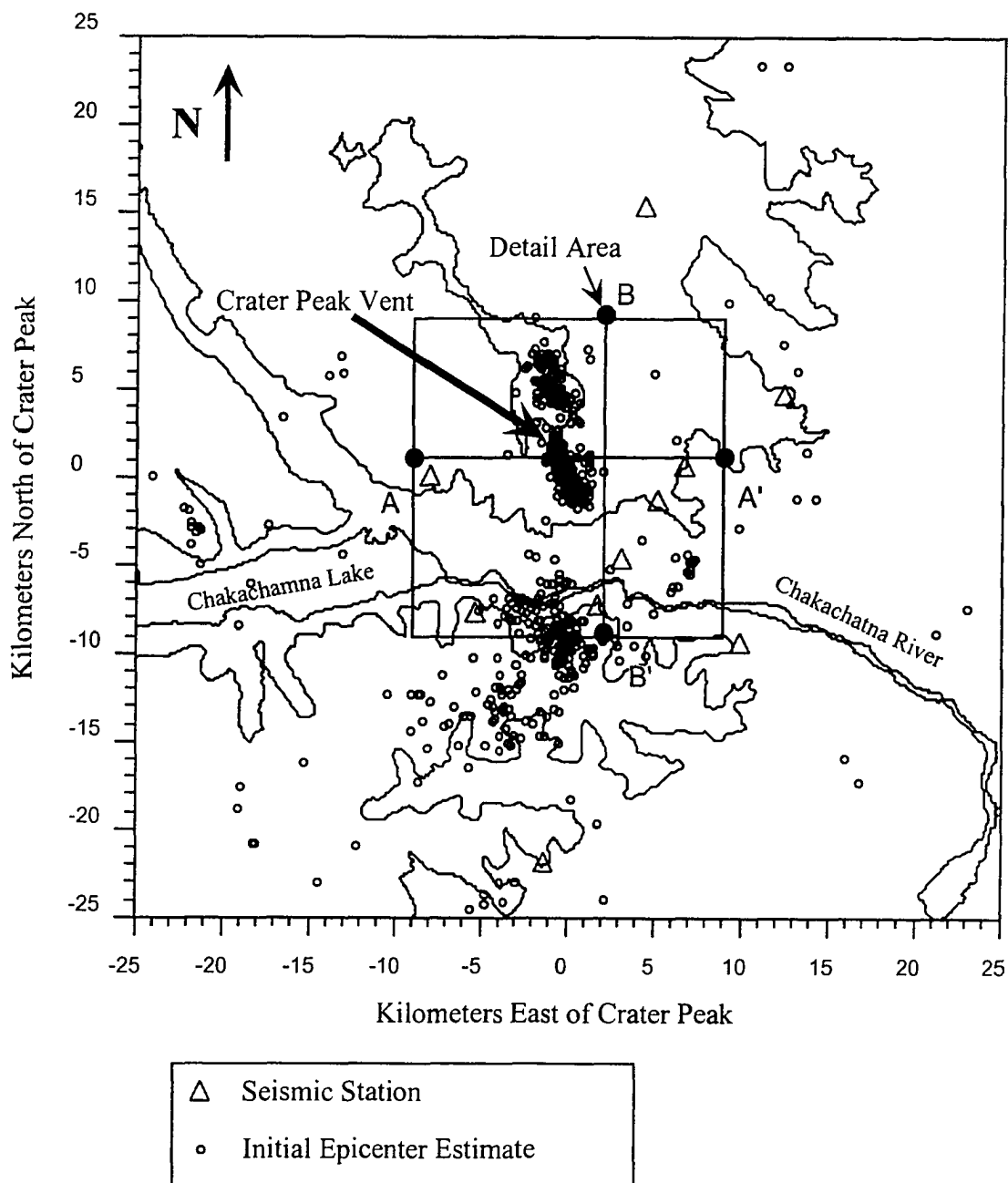


Figure 63: Mt. Spurr study area. The velocity model covers the 18 km x 18 km detail area shown. Initial epicentral estimates for earthquakes used as sources for tomographic imaging are shown as open circles. Seismic stations are shown as open triangles. Topographic contours at 900 and 2750 meters are shown for reference. The location of cross sections A-A' and B-B' are also shown.

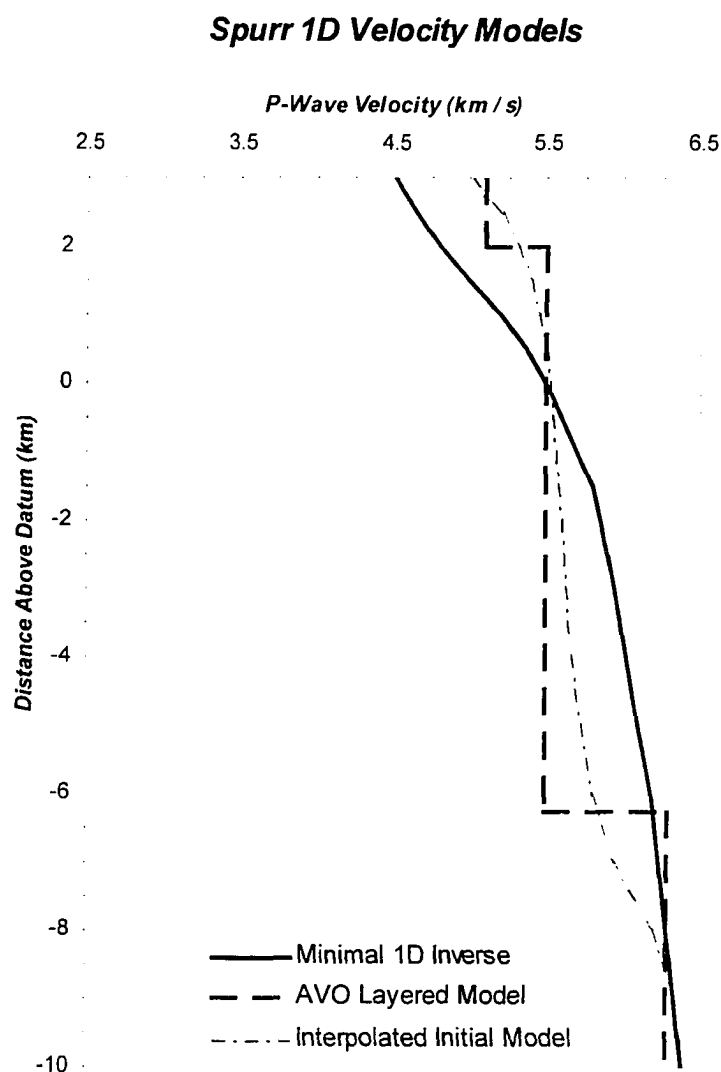


Figure 64: One dimensional Mt. Spurr velocity models.

Tomographic images presented here use the minimal 1D inverse. Data selected as possible imaging sources were relocated with both models. The interpolated model produced 0.020 s larger weighted RMS residual than the minimal 1D inverse. The layered model is used by the Alaska Volcano Observatory for routine earthquake location estimation.

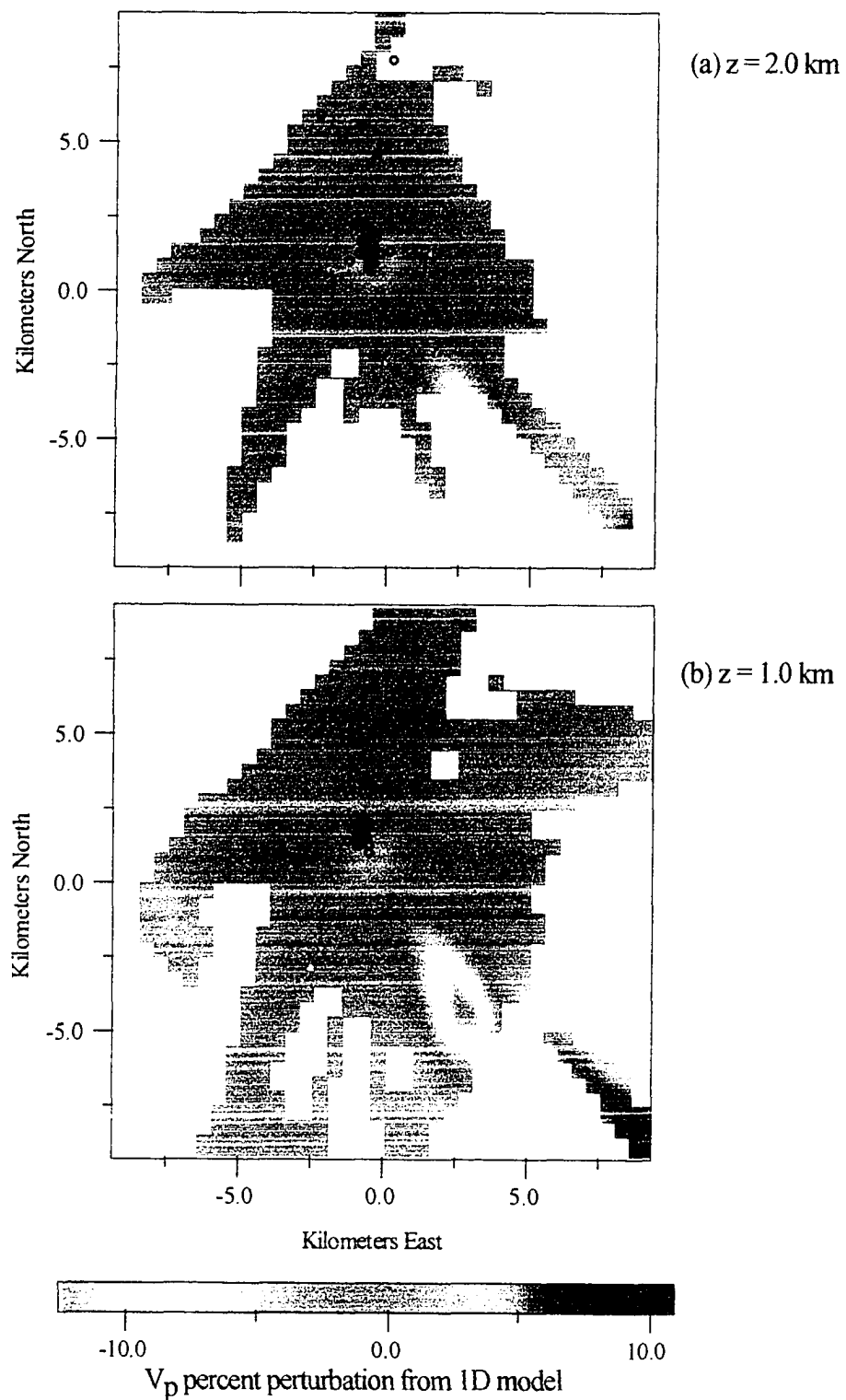


Figure 65: Horizontal slices of P wave velocity perturbation for Mt. Spurr at 2.0 and 1.0 kilometers above sea level. a) 2.0 km above sea level; b) 1.0 km above sea level. Hypocenters within 0.5 km of each plane are shown as open circles.

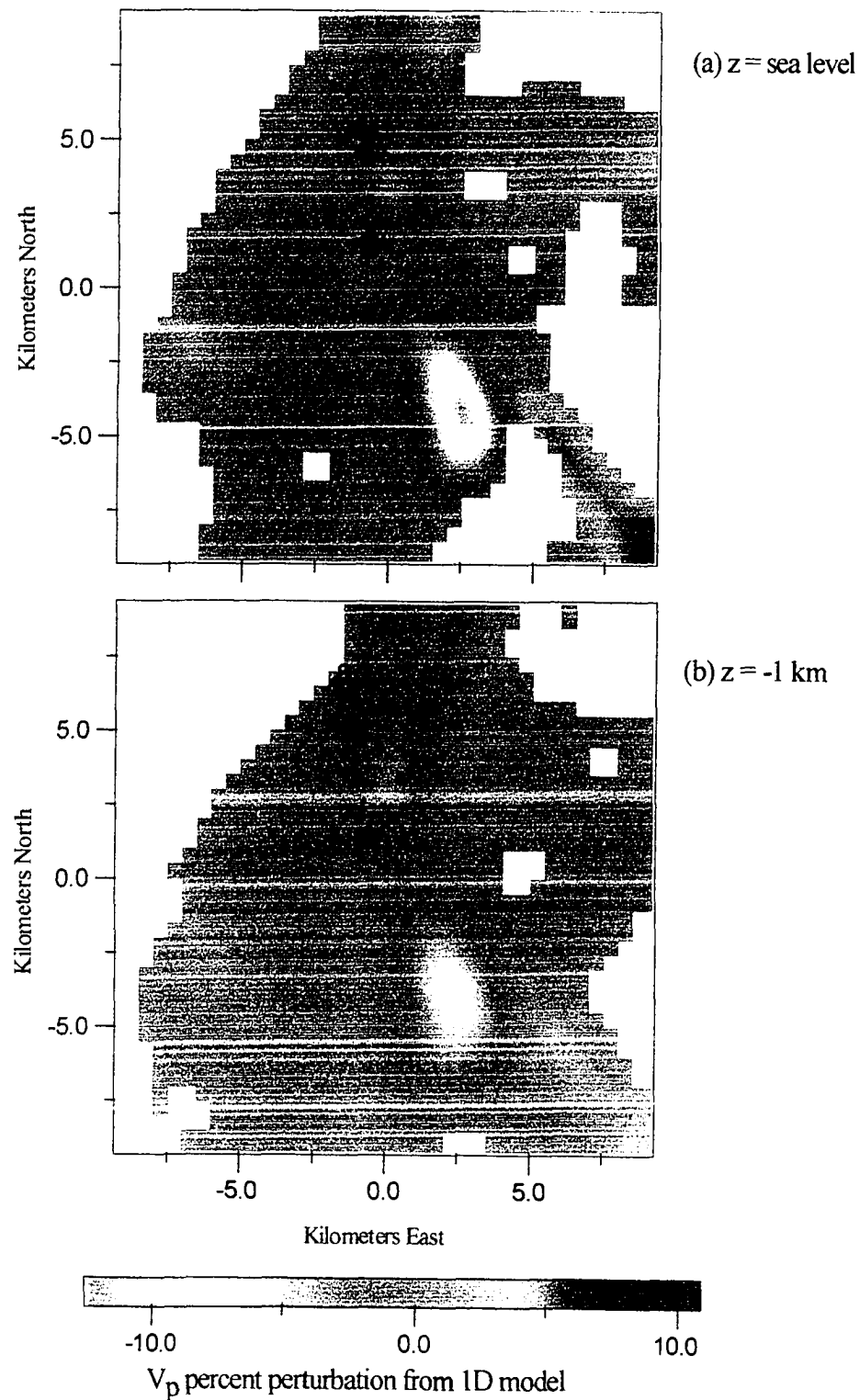


Figure 66: Horizontal slices of P wave velocity estimate for Mt. Spurr at sea level and 1.0 kilometer below sea level. a) Sea level; b) 1.0 km below sea level. Hypocenters within 0.5 km of each plane are shown as open circles.

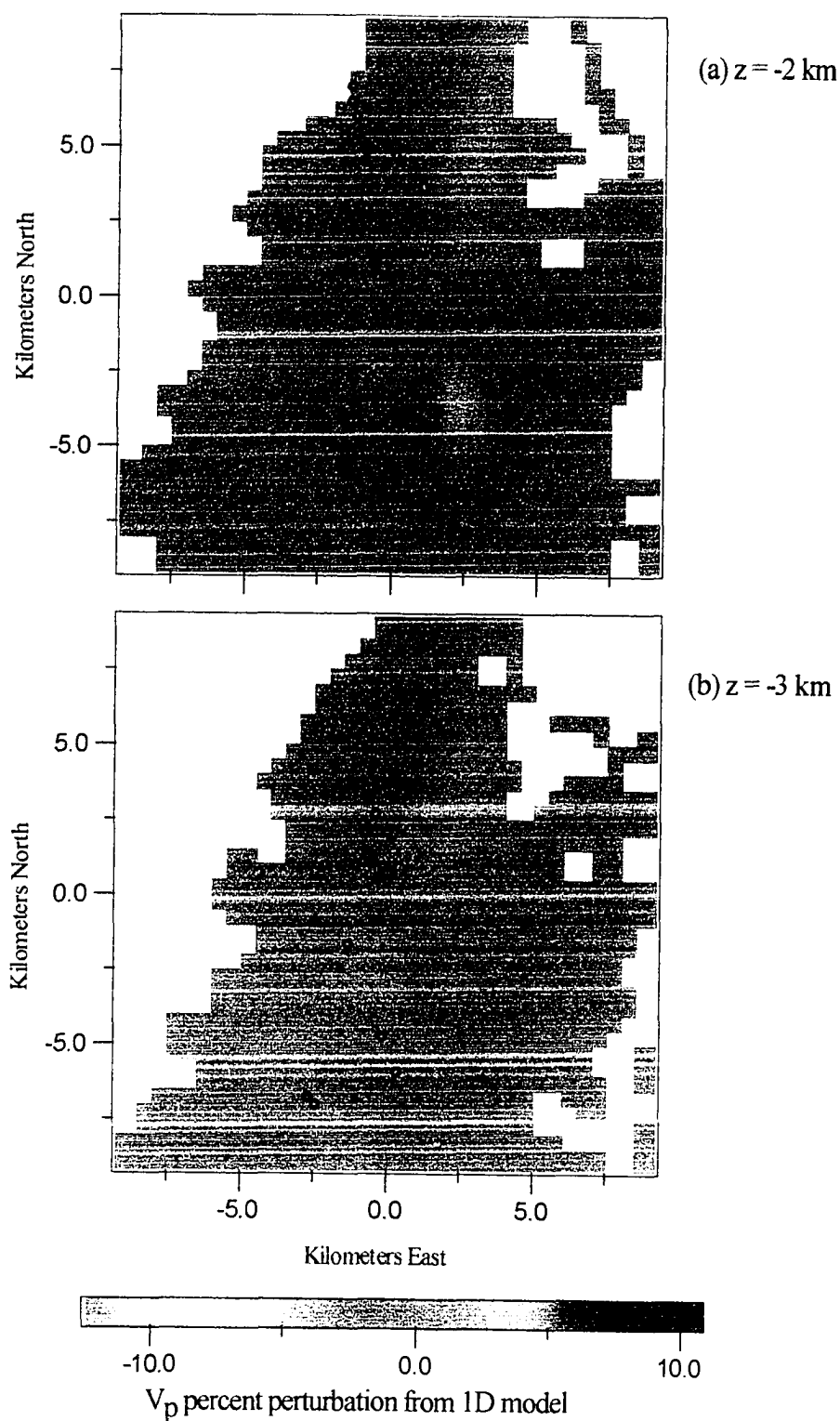


Figure 67: Horizontal slices of P wave velocity perturbation for Mt. Spurr at 2 km and 3 km below sea level. a) 2 km below sea level; b) 3 km below sea level. Hypocenters within 0.5 km of each plane are shown as open circles.

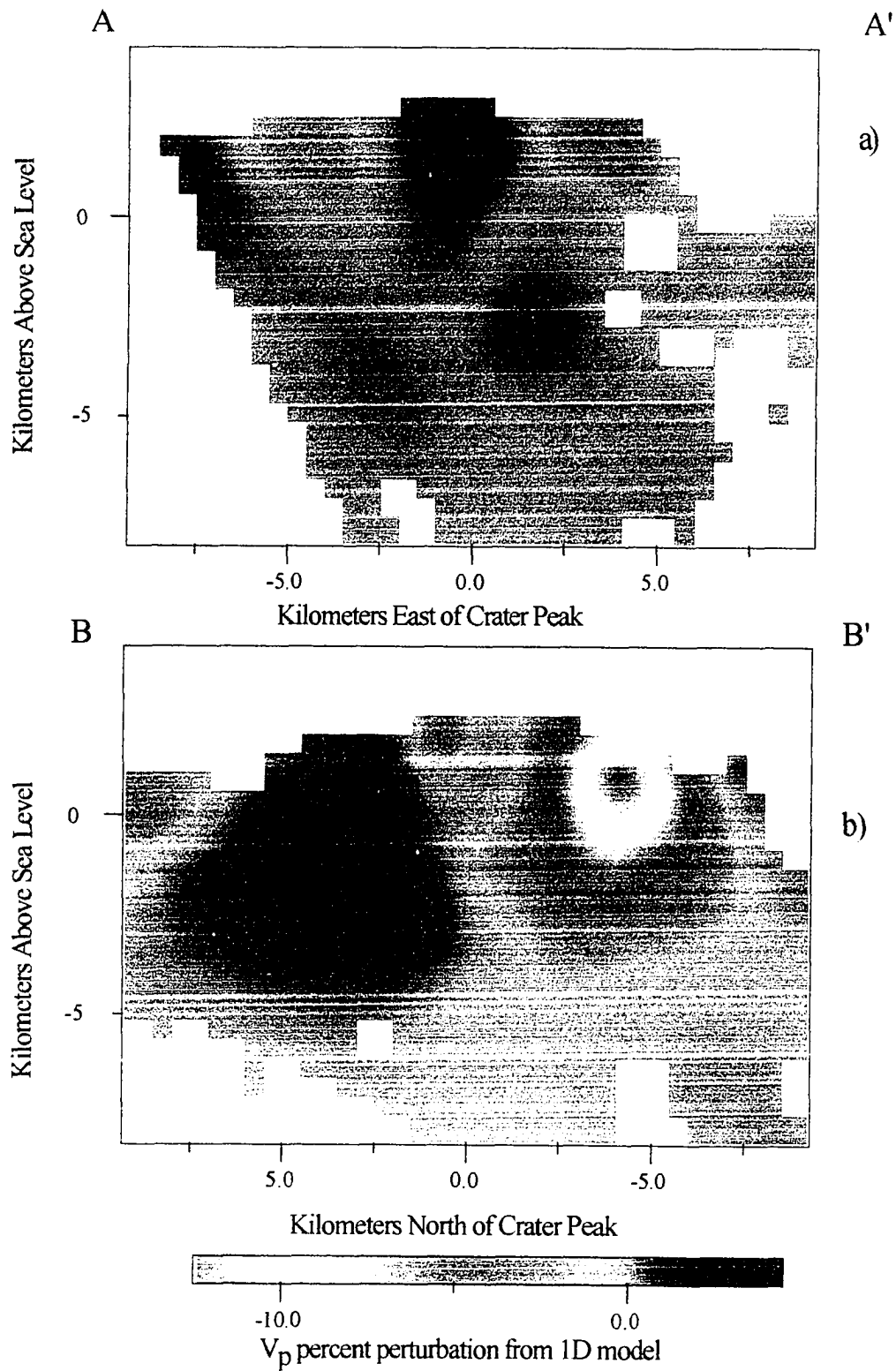


Figure 68: Vertical cross-section of P wave velocity perturbation for Mt. Spurr along profiles A-A' and B-B' shown in Figure 63. a) A-A'; b) B-B'.

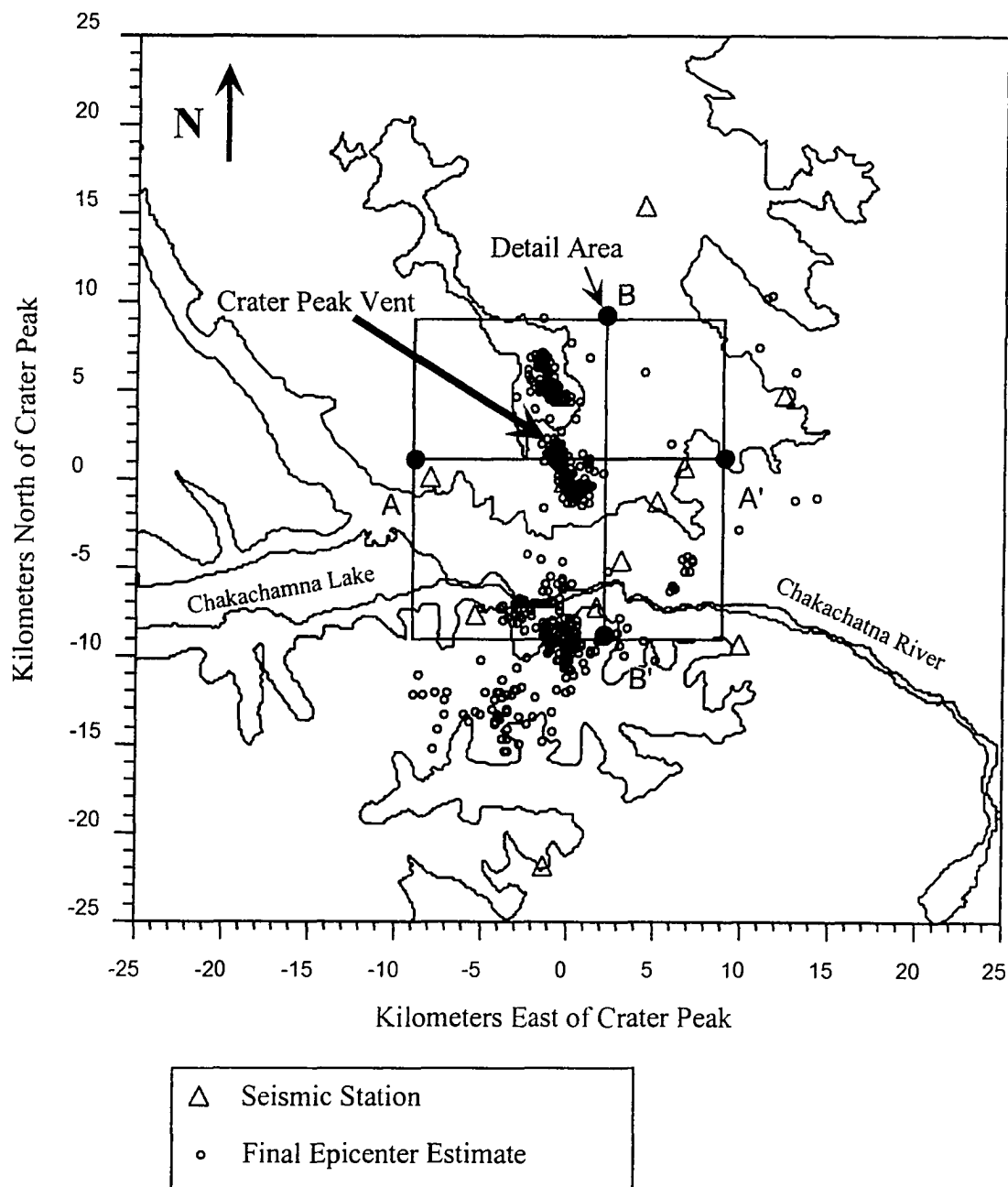


Figure 69: Final relocated hypocenters actually used as sources for tomographic imaging at Mt. Spurr. The velocity model covers the 18 km x 18 km detail area shown. Initial epicentral estimates for earthquakes used as sources for tomographic imaging are shown as open circles. Seismic stations are shown as open triangles. Topographic contours at 900 and 2750 meters are shown for reference. The location of cross sections A-A' and B-B' are also shown.

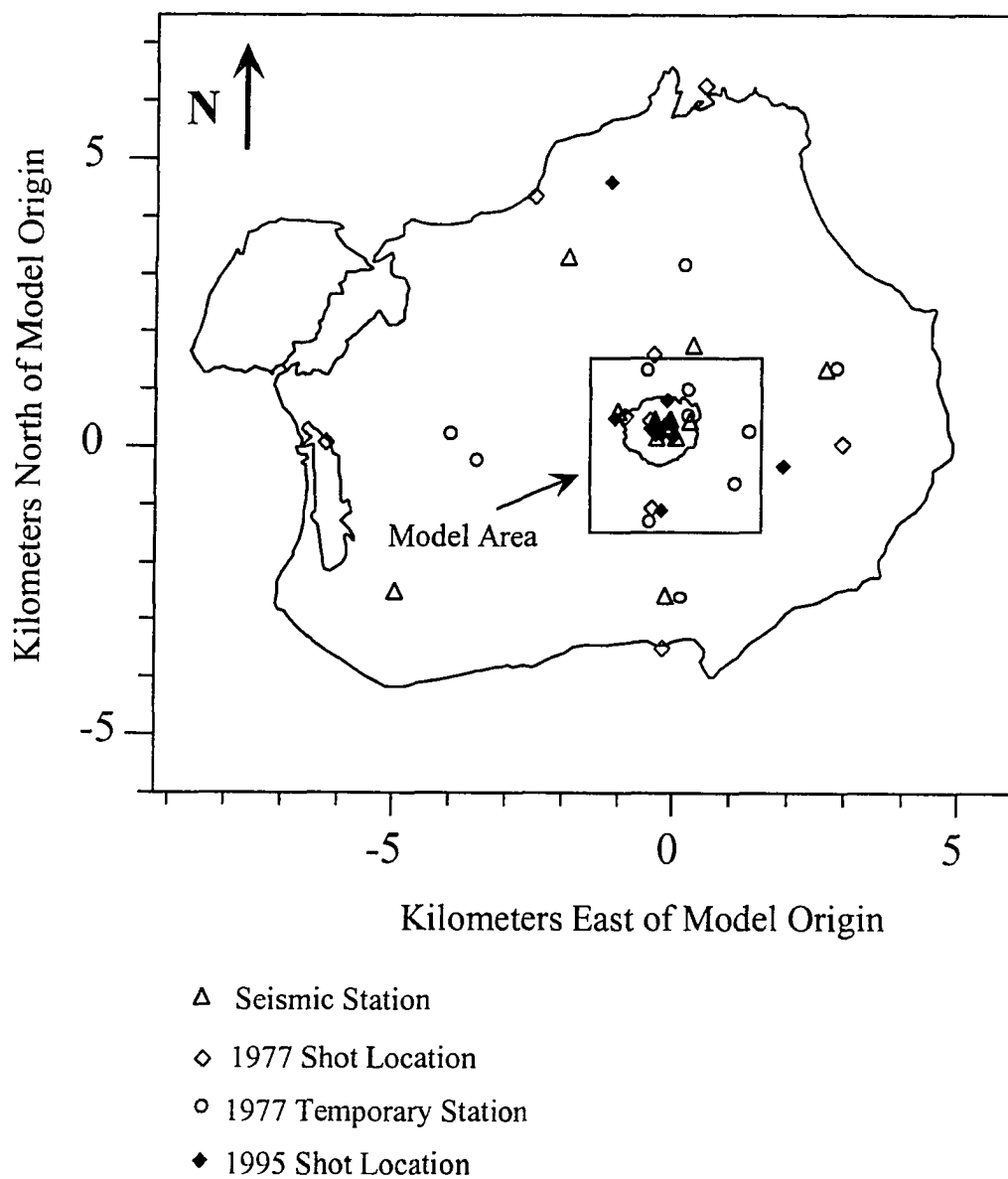


Figure 70: Augustine Volcano study area. The velocity model covers the 3 km x 3 km detail area shown. The outline of the island and the 900 meter elevation contour are shown for reference.

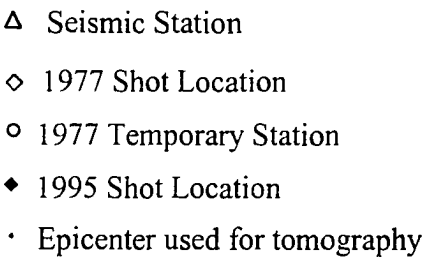


Figure 71: Augustine Volcano velocity model area. The area shown is covered by the tomographic model. The 900 m contour is shown for reference.

Augustine 1D Velocity Models

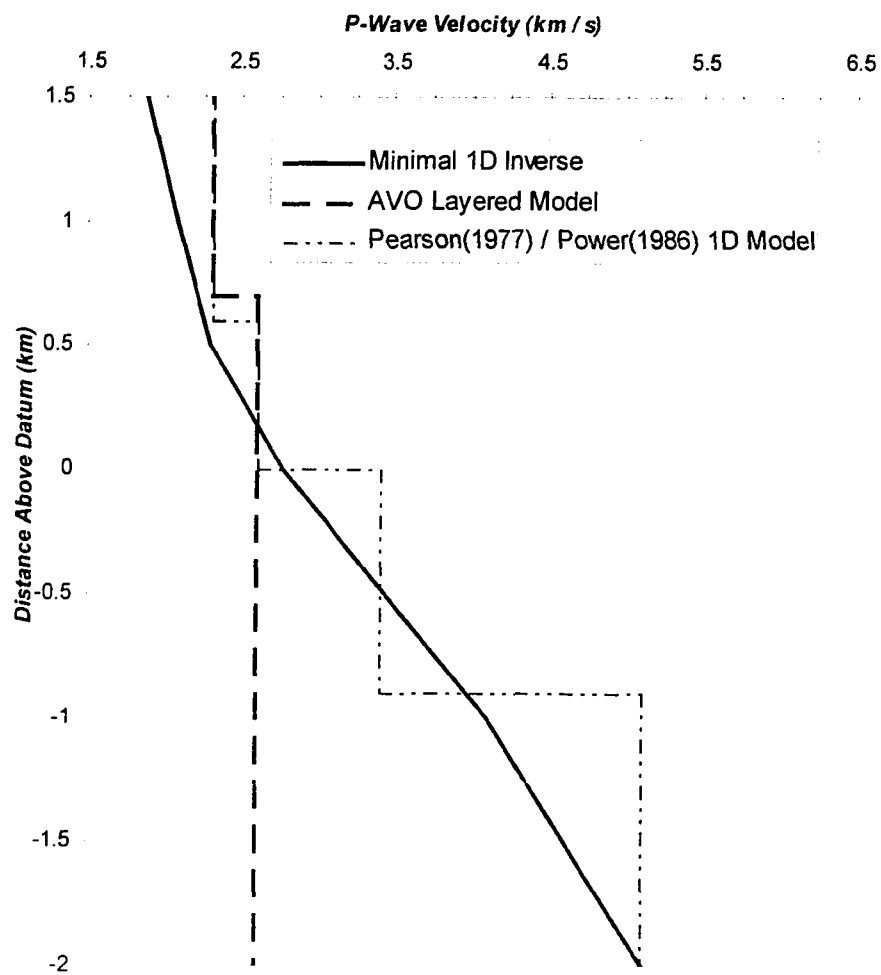


Figure 72: One dimensional Augustine velocity models. Tomographic images presented here use the minimal 1D inverse.

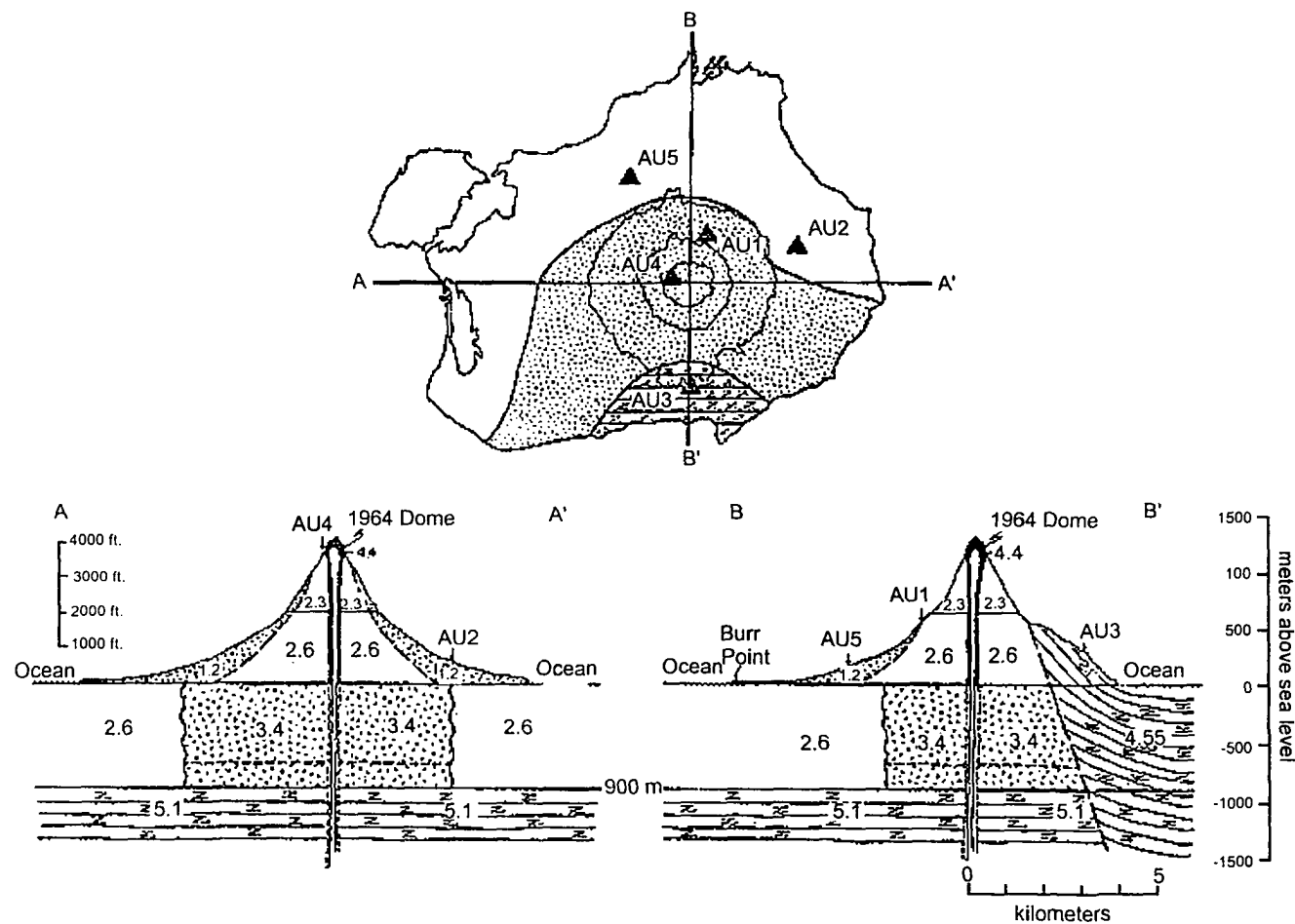


Figure 73: Augustine Volcano velocity model proposed by Kienle et al. (1979). The model is interpretive and extends the 2D model proposed by Pearson (1977). Hypothesized velocities in km s^{-1} are shown for the integrated geological model.

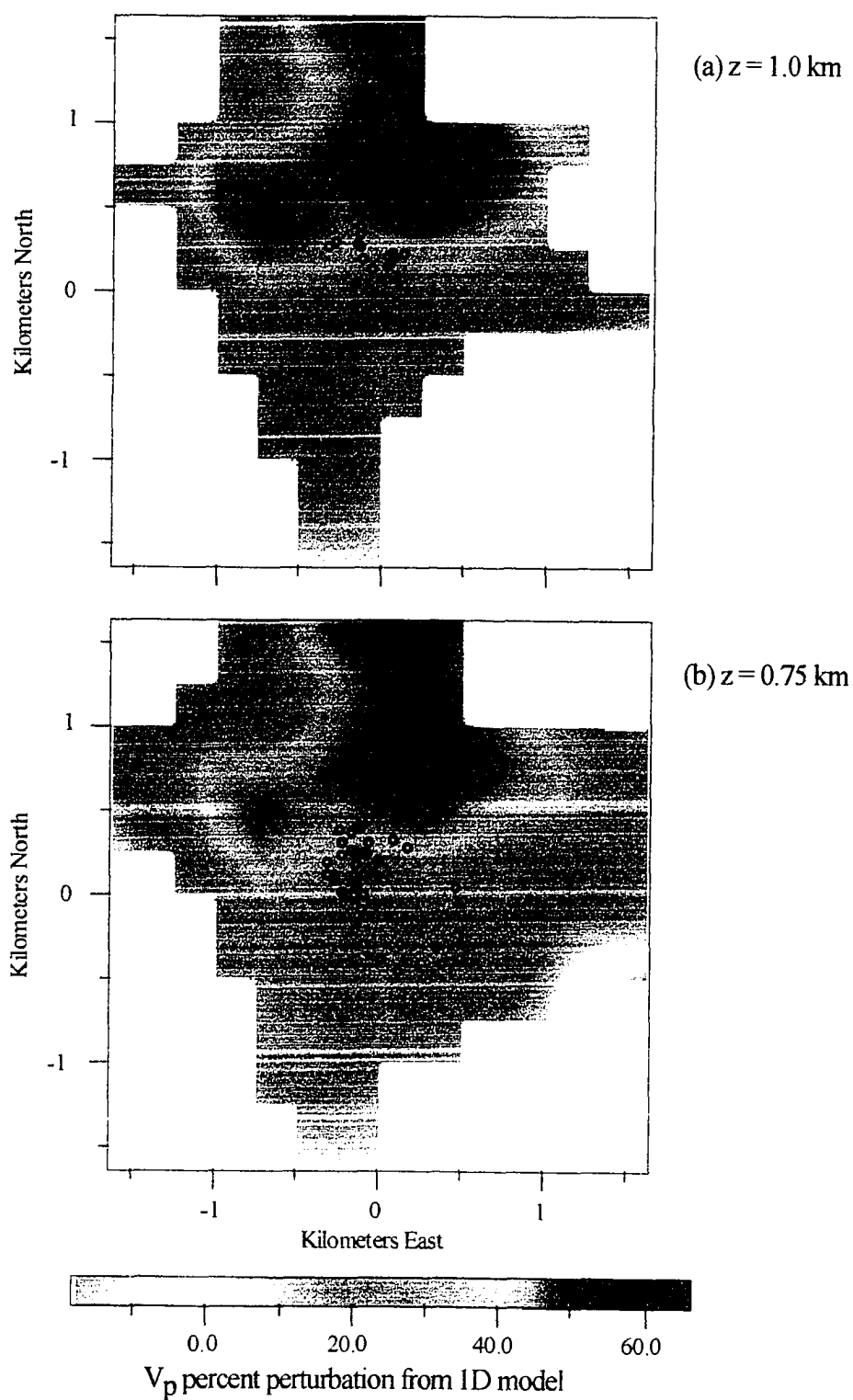


Figure 74: Horizontal slices of P wave velocity perturbation for Augustine Volcano at 1 km and 0.75 km above sea level. a) 1 km above sea level; b) 0.75 km above sea level. Hypocenters within 0.25 km of each plane are shown as open circles.

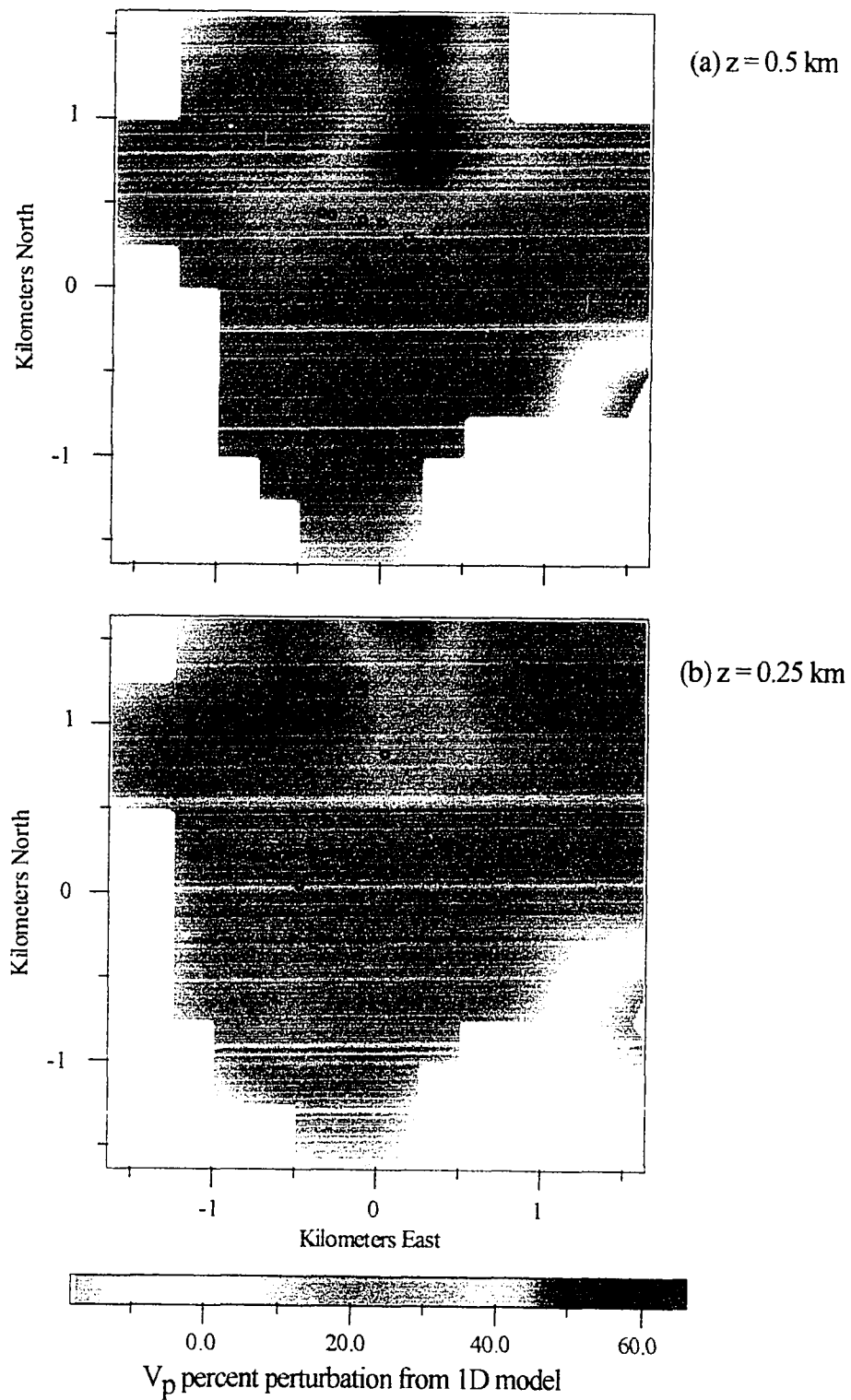


Figure 75: Horizontal slices of P wave velocity perturbation for Augustine Volcano at 0.25 km and 0.5 km above sea level. a) 0.5 km above sea level; b) 0.25 km above sea level. Hypocenters within 0.25 km of each plane are shown as open circles.

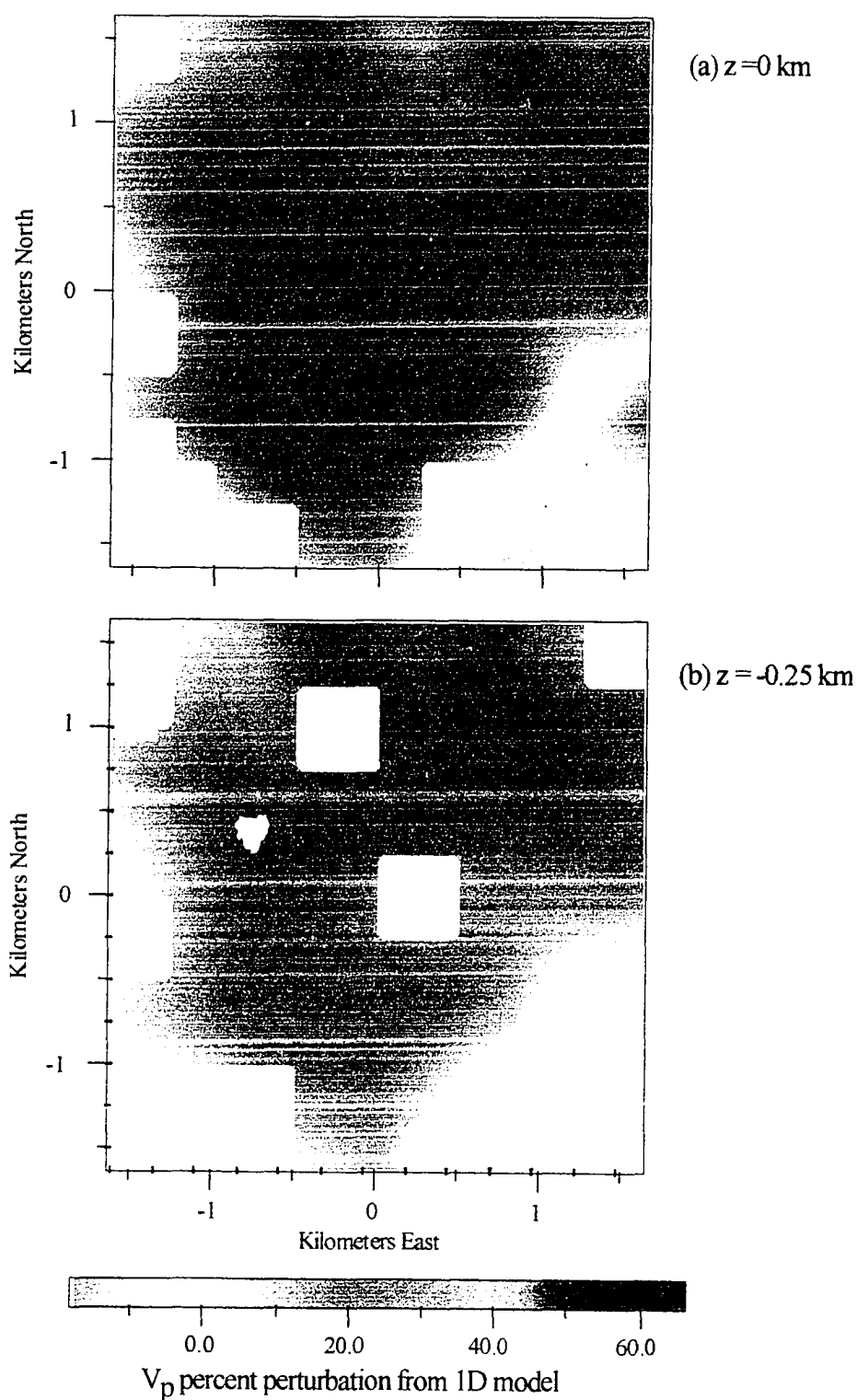


Figure 76: Horizontal slices of P wave velocity perturbation for Augustine Volcano at sea level and 0.25 kilometers below sea level. a) Sea level; b) 0.25 km below sea level. Hypocenters within 0.25 km of each plane are shown as open circles.